

# X-ray cross-correlation analysis applied to disordered two-dimensional systems

R.P. Kurta,<sup>1</sup> M. Altarelli,<sup>2</sup> E. Weckert,<sup>1</sup> and I.A. Vartanyants<sup>1,3,\*</sup>

<sup>1</sup>*Deutsches Elektronen-Synchrotron DESY,  
Notkestraße 85, D-22607 Hamburg, Germany*

<sup>2</sup>*European X-ray Free-Electron Laser Facility,  
Notkestraße 85, D-22607, Hamburg, Germany*

<sup>3</sup>*National Research Nuclear University, “MEPhI”, 115409 Moscow, Russia*

(Dated: February 29, 2012)

## Abstract

Angular x-ray cross-correlation analysis (XCCA) is an approach to study the structure of disordered systems using the results of coherent x-ray scattering experiments. Here, we present the results of simulations that validate our theoretical findings for XCCA obtained in a previous paper [M. Altarelli *et al.*, Phys. Rev. B 82, 104207 (2010)]. We consider as a model two-dimensional (2D) disordered systems composed of non-interacting colloidal clusters with fivefold symmetry and with orientational and positional disorder. We simulate a coherent x-ray scattering in the far field from such disordered systems and perform the angular cross-correlation analysis of calculated diffraction data. The results of our simulations show the relation between the Fourier series representation of the cross-correlation functions (CCFs) and different types of correlations in disordered systems. The dependence of structural information extracted by XCCA on the density of disordered systems and the degree of orientational disorder of clusters is investigated. The statistical nature of the fluctuations of the CCFs in the model ‘single-shot’ experiments is demonstrated and the potential of extracting structural information from the analysis of CCFs averaged over a set of diffraction patterns is discussed. We also demonstrate the effect of partial coherence of x-rays on the results of XCCA.

PACS numbers: 61.05.cp, 61.43.-j, 61.43.Dq, 61.43.Fs

## I. INTRODUCTION

The study of angular correlations in diffraction patterns has a long history with two main directions. On one hand, it has been used in small angle x-ray scattering experiments as a tool in the attempt to solve structures of molecules in solutions or, more generally, in non-periodic systems. It goes back to the work of Kam<sup>1,2</sup>, over 30 years ago; in much more recent times, the perspectives opened by ultra-short and ultra-bright pulses from free-electron lasers<sup>3-6</sup>, which provide an opportunity to acquire diffraction data in a time short compared to translational and rotational relaxation times, have rekindled interest in this approach<sup>7,8</sup>. In these recent applications, the angular correlation is defined as an average over many diffraction patterns.

An alternative purpose, which may seem more modest at first, but could be of paramount importance in the physics of disordered or partially ordered systems is the unveiling of hidden symmetries in a disordered collection of elements. This leads to the problem of understanding systems with the so-called medium range order<sup>9-11</sup> as well as complicated dynamics and correlation between dynamical heterogeneity and medium-range order in a large class of glass-forming liquids<sup>12-16</sup>. In such systems the relevant question will be, for example, can one recognize and identify an  $n$ -fold symmetry axis of an individual molecular species from the diffraction patterns of a liquid composed of such molecules? Can bond angles be detected from the study of angular correlations of the diffracted intensity of an amorphous system? There are partly affirmative answers in the study of partially ordered quasi two-dimensional systems like liquid crystals<sup>17</sup> in which hexatic bond order can be detected by the study of the angular dependence of the diffracted intensity. More recently Wochner *et al.*<sup>18,19</sup> reported angular correlations with pronounced periodic character in a colloidal suspension of polymethylmethacrylate (PMMA) spheres, expected to form icosahedral clusters near the glass formation concentration.

In our first paper<sup>20</sup>, angular correlations in the diffraction patterns from a disordered collection of symmetric species were discussed theoretically in a general frame. The information content of angular correlation data in different experimental conditions (dilute versus dense systems, different statistics of orientational disorder, and so on) was derived. It transpires that generally the richest information is obtainable from dilute systems in which not too many molecules or clusters are found instantly in the illuminated volume. At the same

time for such systems, static fluctuations of the scattered intensity are relatively high, and do not give confirmative information on the local structure of disordered systems. It also appears that partial orientation of the species, with lining up of the symmetry axes in one direction, enhances the effects considerably, in agreement with the results of other groups. Nonetheless, some aspects of the data by Wochner *et al.*<sup>18</sup> were difficult to reconcile with our results, for example the pronounced fivefold symmetry of some of the observed angular correlation patterns.

In this paper we apply the x-ray cross-correlation analysis (XCCA) to the study of the local structure (LS) of disordered systems. In general, the role of LSs can be played by clusters in the gas phase or a colloidal system, protein molecules, viruses, or complex biological systems in solution. Here, we consider the particular case of model 2D systems composed of non-interacting clusters with identical shape. The main goal of this paper is to identify experimental conditions when symmetry and/or structure of clusters composing a disordered system can be correctly determined from the XCCA. By varying the density of clusters and characteristics of their orientational disorder, we analyze how these changes influence the angular cross-correlation functions (CCFs) and their Fourier spectra. We demonstrate a statistical behavior of CCFs calculated for different single realizations of a system and compare them with CCFs averaged over many realizations of the system (ensemble of diffraction patterns). Such analysis can be related to so-called 'single-shot' experiments, when the measurement time is much shorter than any typical structural relaxation time of the system under investigation. Such type of experiments can be performed, for example, on fourth generation x-ray sources, such as free-electron lasers<sup>3-6</sup> (FELs), that produce ultrashort pulses on the femtosecond scale. We also investigate the influence of partial coherence of the incident x-ray beams on the outcome of the XCCA.

This paper is organized as follows. In the next section a brief theoretical description of the XCCA based on our previous work<sup>20</sup> is given. In the third section a model of a coherent scattering experiment on a disordered system used in simulations is introduced. In the fourth section we study the CCF as a function of orientational disorder in the systems with a large number of clusters. Statistical fluctuations of the CCF in a model 'single-shot' experiment for completely disordered systems with a small number of clusters are analyzed in the fifth section. Results of calculations of the averaged CCFs for systems with different number of particles and degree of orientational disorder are presented in the sixth section.

In the seventh section we demonstrate the role of coherence properties of the incoming x-ray wavefields for the analysis of the local structure of disordered systems. The paper is completed by the conclusions and outlook section. Detailed derivations are summarized in the Appendix section.

## II. THEORETICAL BACKGROUND

The intensity CCF  $C_q(\Delta)$  as a function of the angular coordinate  $0 \leq \Delta \leq 2\pi$  is defined as follows<sup>21</sup>

$$C_q(\Delta) = \frac{\langle I(q, \varphi)I(q, \varphi + \Delta) \rangle_\varphi - \langle I(q, \varphi) \rangle_\varphi^2}{\langle I(q, \varphi) \rangle_\varphi^2}, \quad (1)$$

where  $I(q, \varphi)$  is the intensity scattered at the momentum transfer vector  $\mathbf{q}$ ,  $\varphi$  is an angular coordinate around a diffraction ring of radius  $q$ , and  $\langle f(\varphi) \rangle_\varphi = 1/2\pi \int_0^{2\pi} f(\varphi) d\varphi$  denotes the angular average around the ring.

It is convenient to analyze the CCF  $C_q(\Delta)$  using a Fourier series decomposition<sup>20</sup>,

$$C_q(\Delta) = 2 \sum_{n=1}^{\infty} C_q^n \cos(n\Delta), \quad (2a)$$

$$C_q^n = \frac{1}{\pi} \int_0^\pi C_q(\Delta) \cos(n\Delta) d\Delta, \quad (2b)$$

where  $C_q^n$  are the Fourier components.

The Fourier components of the CCF  $C_q(\Delta)$  for  $n \neq 0$  are determined by the angular distribution of intensity  $I(q, \varphi)$  according to the following relation<sup>20</sup>

$$C_q^n = |I^n(q)/I^0(q)|^2, \quad (3)$$

where  $I^n(q) = 1/(2\pi) \int_0^{2\pi} I(q, \varphi) e^{-in\varphi} d\varphi$  are the components of the Fourier expansion of the intensity on the ring of radius  $q$ . According to its definition the zero Fourier component of the intensity distribution  $I^0(q)$  coincides with the small angle x-ray scattering (SAXS) form-factor  $S(q)$ . We also want to note here that normalization in Eq. (3) has to be done with caution, the  $q$ -values where  $I^0(q) = 0$  should be excluded from the consideration.

In the following we will consider a specific case of two-dimensional (2D) disordered systems that on one hand simplifies XCCA analysis but at the same time reveals important general features common for such systems. It was shown<sup>20</sup>, that in this case only even

( $n = 2l, l = 1, 2, 3, \dots$ ) Fourier components of the intensity  $I^n(q)$  are giving non-zero contributions and they can be presented as a sum of two terms, attributed to the different types of structural correlations in the system

$$I^n(q) = (i)^n \left[ \sum_{k=1}^N L_k^n(q) + 2 \cdot \sum_{k_2 > k_1} L_{k_1, k_2}^n(q) \right], \quad n = 2l. \quad (4)$$

Here the first sum over  $k_1 = k_2 = k$  contains the contributions  $L_k^n(q)$  from  $N$  individual clusters, and the second sum over  $k_1 \neq k_2$  contains the cross-terms  $L_{k_1, k_2}^n(q)$  corresponding to correlation of different clusters. In 2D systems these two terms  $L_k^n(q)$  and  $L_{k_1, k_2}^n(q)$  can be presented in the following way<sup>20</sup>

$$L_k^n(q) = \int \int d\mathbf{r}_1 d\mathbf{r}_2 \tilde{\rho}_k(\mathbf{r}_1) \tilde{\rho}_k(\mathbf{r}_2) J_n(q|\mathbf{r}_{21}|) e^{-in\phi_{\mathbf{r}_{21}}}, \quad (5a)$$

$$L_{k_1, k_2}^n(q) = \int \int d\mathbf{r}_1 d\mathbf{r}_2 \tilde{\rho}_{k_1}(\mathbf{r}_1) \tilde{\rho}_{k_2}(\mathbf{r}_2) J_n(q|\mathbf{R}_{k_2, k_1}^{21}|) e^{-in\phi_{\mathbf{R}_{k_2, k_1}^{21}}}. \quad (5b)$$

Here  $\mathbf{R}_{k_2, k_1}^{21} = \mathbf{R}_{k_2, k_1} + \mathbf{r}_{21}$ , where  $\mathbf{R}_{k_2, k_1}$  is the 2D vector connecting centers of different clusters,  $\mathbf{r}_1$  and  $\mathbf{r}_2$  are the 2D vectors defined in the local coordinate system (with its origin located at the center of  $k$ -th cluster). In Eqs. (5a, 5b)  $\phi_{\mathbf{r}_{21}}$  and  $\phi_{\mathbf{R}_{k_2, k_1}^{21}}$  are the angles of the vectors  $\mathbf{r}_{21} = \mathbf{r}_2 - \mathbf{r}_1$  and  $\mathbf{R}_{k_2, k_1}^{21}$  in the external coordinate system,  $\tilde{\rho}_k(\mathbf{r})$  is a projected electron density of a cluster,  $J_n(qr)$  is the Bessel function of the first kind of integer order  $n$ , and the integration is performed over the area of a cluster.

If all clusters in the system have the same internal structure but are randomly oriented in 2D space, the phase  $\phi_{\mathbf{r}_{21}}$  in the exponent of Eq. (5a) can be defined as  $\phi_{\mathbf{r}_{21}} = \phi_k + \phi_{\mathbf{r}_{21}}^0$ , where  $\phi_k$  is the rotation angle of the  $k$ -th cluster with respect to the fixed (reference) angular orientation  $\phi_{\mathbf{r}_{21}}^0$  of the cluster<sup>20</sup>. In this case, the term  $L_k^n(q)$  can be expressed as follows

$$L_k^n(q) = e^{-in\phi_k} L^n(q), \quad (6)$$

and the integral  $L^n(q)$  is the same for all clusters:

$$L^n(q) = \int \int d\mathbf{r}_1 d\mathbf{r}_2 \tilde{\rho}(\mathbf{r}_1) \tilde{\rho}(\mathbf{r}_2) J_n(q|\mathbf{r}_{21}|) e^{-in\phi_{\mathbf{r}_{21}}^0}. \quad (7)$$

It is well seen from the structure of the integral in Eq. (7) that there is a direct correspondence between the electron density  $\tilde{\rho}(\mathbf{r})$  of LSs and the values of  $L^n(q)$ . If  $L^n(q)$  is known, certain information about the electron density  $\tilde{\rho}(\mathbf{r})$ , for example symmetry of clusters, can be obtained. Importantly, the first term in decomposition (4) does not depend

on inter-particle distance and will have the *same* contribution for dilute and dense systems. Contrary to that, according to (5b) the term  $L_{k_1, k_2}^n(q)$  contains the inter-particle distance  $\mathbf{R}_{k_2, k_1}$  as an important parameter. Clearly, the contribution of this term will be different for dilute and dense systems.

In the case of a *dilute* disordered system the contribution of the second sum in Eq. (4) can be neglected<sup>20</sup>, and substituting Eqs. (4) and (6) in Eq. (3) we obtain for the Fourier components  $C_q^n$  of the CCF

$$C_q^n = \mathcal{L}^n(q) \cdot A_n^2, \quad (8)$$

where

$$\mathcal{L}^n(q) = |L^n(q)/L^0(q)|^2, \quad (9)$$

is the structural term, which is directly related to the structure of a LS, and  $A_n$  is the statistical term, which depends on the orientational distribution of LSs in the system. It is defined as an amplitude of the random phasor sum<sup>22,23</sup>,

$$\mathbf{A}_n = A_n e^{i\theta_n} = \frac{1}{N} \sum_{k=1}^N e^{in\phi_k}, \quad (10)$$

where  $\theta_n$  is the phase of this sum. It is clear from Eq. (8) that the value of the Fourier component  $C_q^n$  in dilute systems is determined by the product of the structural and statistical terms.

We analyzed the statistical behavior of  $A_n^2$  for different angular distributions of cluster orientations in our previous paper<sup>20</sup>. Our analysis has shown that in the case of a completely *oriented* dilute system (all  $\phi_k = \phi_0$ ), the statistical term is equal to unity, and the Fourier components reach their maximum values, determined by the structural term  $C_q^n = \mathcal{L}^n(q)$ . If the distribution of orientations of LSs is *Gaussian*, with the standard deviation  $\sigma_\phi$  and a zero mean  $\langle\phi\rangle = 0$ , the average value of the statistical term is

$$\langle A_n^2 \rangle = \exp(-n^2 \sigma_\phi^2) (1 - 1/N) + 1/N. \quad (11)$$

It follows from Eq. (11), that in the limit  $N \rightarrow \infty$ , the Fourier components of the averaged angular CCF are defined by [see for details section VI]

$$\langle C_q^n \rangle = \mathcal{L}^n(q) \exp(-n^2 \sigma_\phi^2). \quad (12)$$

According to these results for a dilute system with Gaussian distribution of orientations, the contribution of higher order Fourier components (corresponding to large  $n$ -values) will be much lower compared to the contribution of the lower order Fourier components.

In the case of a *uniform* distribution of a large number  $N$  of orientations, the statistical term fluctuates around its mean value  $\langle A_n^2 \rangle$  with the standard deviation  $\sigma_{A_n^2}$ :

$$\langle A_n^2 \rangle = 1/N, \quad \sigma_{A_n^2} = 1/N. \quad (13)$$

It means that in this case fluctuations are of the same order as the mean value and their ratio does not depend on the number of clusters in the system. According to Eqs. (13) and (8), in the limit of a large number of orientations ( $N \rightarrow \infty$ ) the Fourier components of the averaged CCF have vanishing values,  $\langle C_q^n \rangle \rightarrow 0$ .

In the case of a *dense* disordered system, when the average distance between LSs is of the order of the size of a single cluster, the second sum in Eq. (4) can not be neglected. Importantly, for such a dense system it can significantly affect the Fourier spectrum of the angular CCF. Taking into account both terms of Eq. (4), the Fourier components of the angular CCF (3) can be written as the following sum<sup>24</sup>

$$C_q^n = \left| \frac{I^n(q)}{I^0(q)} \right|^2 = \frac{S_1^n + (S_2^n + S_3^n) + S_4^n}{|I^0(q)|^2}, \quad (14)$$

where

$$S_1^n = \left| \sum_k L_k^n(q) \right|^2 = N^2 |L^n(q)|^2 A_n^2, \quad (15a)$$

$$S_2^n + S_3^n = 4 \cdot \text{Re} \left[ \sum_k L_k^{n*}(q) \cdot \sum_{k_2 > k_1} L_{k_1, k_2}^n(q) \right], \quad (15b)$$

$$S_4^n = 4 \left| \sum_{k_2 > k_1} L_{k_1, k_2}^n(q) \right|^2, \quad (15c)$$

and

$$|I^0(q)|^2 = \left| N L^0(q) + 2 \sum_{k_2 > k_1} L_{k_1, k_2}^0(q) \right|^2. \quad (16)$$

The summation in Eqs. (15a-15c) is performed over all LSs. In Eq. (14) the term  $S_1^n$  is defined by the local structure, and terms  $S_2^n$ ,  $S_3^n$  and  $S_4^n$  contain contributions from the inter-particle correlations due to the second term in Eq. (4). Below we will explicitly investigate contribution of these terms to the total value of the Fourier components  $C_q^n$  of the CCF.

### III. MODEL OF A DISORDERED SYSTEM AND SCATTERING GEOMETRY

#### (a) *Model*

The model and the geometry of the coherent x-ray scattering experiment used in our simulations<sup>25</sup> is shown in Fig. 1(a). A coherent x-ray beam with a wavelength  $\lambda = 1.5 \text{ \AA}$  kinematically scatters from a 2D disordered sample of the total area  $10 \times 10 \text{ } \mu\text{m}^2$ . Diffraction patterns are recorded on a 2D detector of the size  $D = 12 \text{ mm}$  (with the pixel size  $p = 10 \text{ } \mu\text{m}$ ), positioned in the far-field at the distance  $L = 2.5 \text{ m}$  from the sample. The chosen scattering geometry allows to cover a  $q$ -range up to  $0.1 \text{ nm}^{-1}$  (that corresponds the maximum scattering angle,  $\theta_{\text{max}} = 0.14^\circ$ ), and also corresponds to about four pixels per speckle in a diffraction pattern for a given sample size<sup>26</sup>.

As a model of a disordered sample a 2D system composed of identical clusters, generally in random positions and orientations, is considered [Fig. 1(b)]. We chose a centered pentagonal 2D cluster with a 5-fold rotational symmetry for our simulations [see the inset in Fig. 1(b)]. This symmetry belongs to the forbidden motifs in the long-range crystalline matter but may exist on the short length scales in disordered systems<sup>27</sup>, and is, clearly, most intriguing for a detailed study. Each cluster is composed of close-packed colloidal spheres with a radius of  $100 \text{ nm}$  giving the total cluster size  $d = 600 \text{ nm}$ . The sample is oriented with its 2D plane perpendicular to the direction of the incident x-ray beam [see Fig. 1(a)], and the 5-fold rotational axis of each pentagonal cluster in the sample is parallel to the direction of the incident x-ray beam. The choice of parameters for our model was selected in analogy to the system studied experimentally in Ref.<sup>18</sup>.

In our simulations disordered samples are generated in the following way. Starting with an ordered 2D lattice of clusters, positional disorder is achieved by a sequence of random movements of all clusters. A single randomization step consists of Voronoi tessellation<sup>28</sup> of the sample area, with each Voronoi cell containing only one cluster, and a subsequent displacement of each cluster by a random vector within the Voronoi cell. After  $10 - 50$  steps there is no positional order in the system. This procedure is repeated several times to generate the sequence of uncorrelated systems. Orientational disorder of clusters is implemented on the final step by applying a certain angular distribution of clusters to each of the generated samples. In Fig. 1(b) one of the typical realizations of this procedure is presented.



(b) *Structural term*

As discussed earlier, the contribution of a single cluster to different Fourier components of the CCF is determined by the term  $\mathcal{L}^n(q)$  in Eq. (8) and strongly depends on its symmetry and structure. In case of a centered pentagonal cluster considered in our simulations due to its 5-fold symmetry only Fourier components with  $n = 10l$ , where  $l$  is an integer number, have a non-zero contribution (see Appendix A). Importantly, the same term in Eq. (8) determines the  $q$ -dependence of Fourier components  $C_q^n$  of the CCF. In Fig. 2 the values of  $|L^n(q)|^2$  and  $|L^0(q)|^2$  normalized by the form-factor of a sphere  $|f(q)|^2$  as well as their ratio  $\mathcal{L}^n(q) = |L^n(q)/L^0(q)|^2$  as a function of  $q$  are presented for  $n = 10, 20$  and  $30$ . As a general rule, at low  $q$ -values, due to the properties of the Bessel functions, the strongest contribution to the CCF comes from the Fourier component with the lowest  $n$  ( $n = 10$  in our case). At higher  $q$ -values contribution of other components becomes comparable. Notice, that at some  $q$ -values the function  $\mathcal{L}^n(q)$  contain sharp maxima corresponding to  $q$ -values where  $|L^0(q)|^2$  has small values. In a dilute system where we can neglect inter-particle correlations the factor  $NL^0(q)$  coincides with the structure factor  $S(q)$  in SAXS. It means that one can expect an enhancement of the structural term for small  $S(q)$  values. Similar observations were made in experiment<sup>18</sup>. Examination of Fig. 2(c) also shows that small variations of  $q$ -values can result in large variations in the values of  $\mathcal{L}^n(q)$  and consequently in  $C_q^n$ . This is also similar to results obtained in Ref.<sup>18</sup>. It can be also seen in Fig. 2(c) that there are some regions in  $q$  where the contributions of all components are close to zero. We may expect, that at these specific  $q$ -values the contribution of other terms in Eq. (14), that are responsible for inter-cluster correlations in dense systems, can become significant.

The values of the structural term  $\mathcal{L}^n(q)$  at five selected values of  $q$  [shown by arrows in Fig. 2(c)] are presented in Table I. In the following we will compare these values with the values obtained for different systems by means of the XCCA.

#### IV. LARGE NUMBER OF CLUSTERS IN THE SYSTEM

In this section we consider model systems with different characteristics of orientational disorder. We are specifically interested here in analyzing the case of a dense system, contrary to the analysis of the dilute systems performed in our previous paper<sup>20</sup>. In the following we consider three systems consisting of 121 pentagonal clusters<sup>29</sup> with the same positions

of clusters in all three systems but with different orientational disorder in each system [see Fig. 1(b)]. These three systems are: a completely oriented system of clusters (Fig. 3), a partially oriented system described by the Gaussian distribution of orientations with the standard deviation  $\sigma_\phi = 0.05 \cdot 360^\circ/5 = 3.6^\circ$  and zero mean (Fig. 4) and one with a uniform distribution of orientations in the angular range<sup>30</sup>  $(-\pi/5, \pi/5)$  (Fig. 5). The relative separation between the clusters in all three cases is characterized by the ratio  $\langle R \rangle/d \approx 1.5$ , where  $\langle R \rangle$  is the average distance between the clusters estimated from  $\langle R \rangle \sim 1/\sqrt{n_0}$ , and  $n_0$  is the number of clusters per unit area (number density).

Diffraction patterns for each case are presented in Figs. 3(a), 4(a) and 5(a). The most prominent features for all diffraction patterns are concentric rings appearing due to the contribution of the form-factor  $f(q)$  of a single colloidal sphere to the scattered intensity. The speckles superimposed on the diffraction patterns originate from the interference of x-rays coherently scattered on the spatially disordered clusters, distributed within a finite sample area. The corresponding angular averaged intensities  $\langle I(q, \varphi) \rangle_\varphi$  for each diffraction pattern, as well as the angular distribution of orientations of clusters in each system are presented in Figs. 3(b), 4(b) and 5(b). The following momentum transfer  $q$ -values:  $q_1 = 0.023 \text{ nm}^{-1}$ ,  $q_2 = 0.029 \text{ nm}^{-1}$ ,  $q_3 = 0.036 \text{ nm}^{-1}$ ,  $q_4 = 0.043 \text{ nm}^{-1}$  and  $q_5 = 0.059 \text{ nm}^{-1}$  were considered for the detailed analysis in each case [shown in Figs. 3(b), 4(b) and 5(b) by red circles]. The CCFs  $C_q(\Delta)$  as well as corresponding Fourier components  $C_q^m$  were evaluated using Eqs. (1, 2b) for each system at these selected  $q$ -values and are presented in Figs. 3(c,d), 4(c,d) and 5(c,d).

Comparison of all three cases shows that similar to a dilute system the dominant Fourier components ( $n = 10$  and  $20$  in our case) are well defined for more oriented systems (Fig. 3 and Fig. 4). It is also well seen that for a Gaussian distribution of orientations (Fig. 4) at higher  $q$ -values ( $q = q_5$  in this case) the magnitudes of the Fourier components with  $n = 10$  and  $n = 20$  are significantly reduced in comparison to a completely oriented system (Fig. 3). This decay can be attributed to the exponential factor present in the statistical term in Eqs. (11,12).

Our analysis of the first two systems with the high degree of orientational order shows that the symmetry of clusters composing the system can be well determined by means of the XCCA analysis. However, the absolute values of Fourier components can be strongly affected by the inter-particle correlations (see Tables I, II). Comparison of the absolute values

of the Fourier components  $C_q^n$  at  $n = 10, 20$  for the completely oriented system considered here shows that they may fluctuate around the value of the structural term  $\mathcal{L}^n(q)$  for about few tens of percent at different  $q$ -values (see Table I). Results of similar calculations for a dense system with the Gaussian distribution of orientations are presented in Table II. They are compared with the averaged Fourier components of CCFs defined in Eq. (12). They show similar fluctuations of the values of  $C_q^n$  due to inter-particle correlations in the system and statistical fluctuations of orientational distribution of clusters (see next section for a detailed discussion).

Our simulations for the systems with different positional disorder have shown that  $C_q^n$  also depend on the particular realization of the system. Such behaviour of  $C_q^n$  can be explained by the specific constructive and destructive interference of the scattered x-rays in each particular system. Due to the interference term  $(S_2^n + S_3^n)$  in Eq. (14) the inter-particle contribution can increase or decrease the contribution of the structural term  $S_1^n$  in a dense system. Note, that the term  $S_4^n$  is always non-negative and can only increase the value of  $C_q^n$ . That brings us to the conclusion that determination of the absolute values of the Fourier coefficients  $C_q^n$ , or  $|I^n(q)|$  that is necessary for the direct structural determination of the clusters by phase retrieval<sup>8</sup> can be strongly affected in close packed systems.

Contrary to these two cases, the system with the uniform distribution of orientations shows quite different behavior. As a result of the complete orientational disorder, the CCFs shown in Fig. 5(c) do not represent a simple  $\sim \cos(n\Delta)$  form but have rather complicated angular dependence that can be also well observed in the corresponding Fourier spectra in Fig. 5(d). For example, the component with  $n = 8$  (and some other higher values of  $n$ ) dominates at  $q = q_1$  contrary to the previous more ordered systems, where the component with  $n = 10$  was strongly dominant at that  $q$ -value. It is well seen in Fig. 5(d) that in this case there are many Fourier components with comparable values in a wide range of the spectrum. We should note also here that the absolute values of all Fourier components in this case are about one order of magnitude lower then in the case of the ordered systems [see Figs. 3(d) and 4(d)]. This general behavior of disordered systems can be also explained by examining the contribution of two terms in Eq. (4). In the system with the uniform distribution of angular orientations the absolute value of the first term in Eq. (4) is significantly reduced due to the asymptotic behavior of the statistical term  $A_n^2$  given in Eq. (13). For a large number of orientations it drops as  $1/N$  and becomes much smaller then the second

term in Eq. (4) that is determined mostly by the inter-particle correlations. In this case inter-particle correlations give the major contribution to CCFs and determination of the structure or symmetry of the clusters composing the system becomes practically impossible.

The results of our simulations for the 2D samples with different degree of orientational disorder presented in this section clearly indicate, that the possibility to extract the information on the symmetry of individual clusters in the systems with a large number of clusters by the cross-correlation analysis is strongly related to the degree of orientational disorder in the system. The preferred situation is when the system is partially aligned, while correlation analysis is practically impossible in a completely disordered dense system.

## V. ENSEMBLE-DEPENDENT FLUCTUATIONS OF CORRELATIONS

In this section we will consider in more detail the statistical behavior of the CCFs as a function of orientational disorder in a system with a small number of possible orientations. As was discussed above, in a dilute disordered system composed of a certain number of clusters the statistical term  $A_n^2$  in the CCF is fluctuating around an average value  $\langle A_n^2 \rangle$  [see Eqs. (11) and (13)], determined by the statistics (Gaussian, uniform, *etc.*) and number  $N$  of angular orientations of clusters. Contrary to the previous section where we analyzed systems with a large number of clusters ( $N = 121$ ), here we consider systems with much smaller number of clusters for which statistical fluctuations of the scattered intensity become considerably larger.

We analyze here three samples, each composed of  $N = 11$  clusters. The positions of clusters are the same in each sample, and their relative separation is given by the ratio  $\langle R \rangle / d \sim 5.0$ . We consider for each sample a uniform distribution of 11 orientations of clusters within the angular range  $(-\pi/5, \pi/5)$  [see Fig. 6].

The results of the Fourier analysis of the CCFs for these three systems at the selected values of  $q$  are shown in Fig. 6 (note that a different vertical scale for the Fourier spectra was used at different  $q$  values). Comparing the Fourier spectra calculated at the same  $q$  value for these systems one can see, that they significantly vary only due to a different orientational distribution of clusters from one sample to another. In the case of dilute systems, fluctuations of the values of Fourier components are defined by the statistical distribution of orientations of clusters [see Eqs. (8-13)]. In the systems considered here, these fluctuations are defined

by the interplay of two factors, the distribution of cluster orientations [term  $S_1^n$  in expansion (14)], and their relative spatial positions (terms  $S_2^n$ ,  $S_3^n$  and  $S_4^n$ ). This leads in some cases to a large contribution of  $S_1^n$ , sufficient to dominate over the other terms in Eq. (14) [see, for example, the results for the Sample 3 in Fig. 6(c)], and in some cases to a smaller contribution, allowing the inter-cluster correlations to dominate in the Fourier component  $C_q^n$  [see, for example, the results for the Sample 1 in Fig. 6(a)]. Therefore, even if the number of possible orientations of clusters in a system is small, it may be not possible to extract the information on local structure of such a system from a single diffraction pattern. The statistical behavior of CCFs discussed here could give a possible explanation to experimental observations in Ref.<sup>18</sup>, where diffraction patterns taken at different times show different types of angular correlations.

As demonstrated here in a system with a small number of randomly oriented clusters reliable information on the local structure and symmetry of the clusters can not be obtained by the analysis of a single diffraction pattern due to statistical fluctuations of intensity. We discuss in the next section how this problem can be solved in certain cases by averaging of CCFs.

## VI. EFFECTS OF ORIENTATIONAL DISORDER ON THE AVERAGED CCF

In our previous discussion we presented the results of the XCCA applied to the analysis of the diffraction data obtained for a single realization of a disordered system. We showed that statistical fluctuations of CCFs from one realization to another is a strong limitation for the determination of reliable information on cluster symmetry and structure. To overcome this difficulty it was proposed, first by Kam<sup>1,2</sup>, to analyze CCFs averaged over an ensemble of diffraction patterns, instead of analyzing CCFs calculated for a single diffraction pattern (see also recent publications<sup>7,8,31</sup>). Below we will investigate conditions when this approach is valid.

We will consider here the CCF  $\langle C_q(\Delta) \rangle_M$  averaged over a sufficiently large number  $M$  of diffraction patterns,

$$\langle C_q(\Delta) \rangle_M = 1/M \sum_{m=1}^M C_q^m(\Delta), \quad (17)$$

where  $C_q^m(\Delta)$  is the CCF obtained for the  $m$ -th diffraction pattern [Eq. (1)]. Performing the

Fourier transform of both parts of Eq. (17) and taking into account that Fourier transform is a linear operator we obtain for the Fourier components of the averaged CCFs

$$\langle C_q^n \rangle_M = 1/M \sum_{m=1}^M \{C_q^n\}^m, \quad (18)$$

where  $\{C_q^n\}^m$  are the  $n$ -th Fourier components of the  $m$ -th CCF  $C_q^m(\Delta)$ . Eq. (18) means, that the  $n$ -th Fourier component of the averaged CCF is an average of the corresponding Fourier components determined for each diffraction pattern. Taking into account expression (14) for each individual Fourier component of the CCF we obtain

$$\langle C_q^n \rangle_M = \left\langle \left| \frac{I^n(q)}{I^0(q)} \right|^2 \right\rangle_M = \left\langle \frac{S_1^n}{|I^0(q)|^2} \right\rangle_M + \left\langle \frac{(S_2^n + S_3^n)}{|I^0(q)|^2} \right\rangle_M + \left\langle \frac{S_4^n}{|I^0(q)|^2} \right\rangle_M, \quad (19)$$

where  $S_j^n$  are defined by Eqs. (15a-15c) and  $|I^0(q)|^2$  by Eq. (16). Below we show by direct simulations that for a sufficiently large number of diffraction patterns  $M$  the ensemble averaged term  $\langle (S_2^n + S_3^n)/|I^0(q)|^2 \rangle_M$  asymptotically approaches zero and we have for the Fourier components of the averaged CCF

$$\langle C_q^n \rangle_M = \langle C_{q \text{ (clust)}}^n \rangle_M + \langle C_{q \text{ (int-clust)}}^n \rangle_M, \quad (20)$$

where

$$\langle C_{q \text{ (clust)}}^n \rangle_M = \langle S_1^n / |I^0(q)|^2 \rangle_M, \quad (21a)$$

$$\langle C_{q \text{ (int-clust)}}^n \rangle_M = \langle S_4^n / |I^0(q)|^2 \rangle_M. \quad (21b)$$

This result indicates that in a general case for a sufficiently large number of diffraction patterns the Fourier components of the averaged CCF can be represented as an additive sum of two positive valued contributions. The first one is given by the structure of individual clusters and the second one is determined by the inter-cluster contribution.

In the case of a dilute disordered system, when the inter-cluster contribution  $\langle C_{q \text{ (int-clust)}}^n \rangle_M$  in Eq. (20) can be neglected, we obtain

$$\langle C_q^n \rangle_M = \langle C_{q \text{ (clust)}}^n \rangle_M = \mathcal{L}^n(q) \cdot \langle A_n^2 \rangle_M, \quad (22)$$

where  $\langle A_n^2 \rangle_M$  is an ensemble averaged square amplitude of the random phasor sum (10)

$$\langle A_n^2 \rangle_M = 1/M \sum_{m=1}^M \{A_n^2\}^m. \quad (23)$$

For a sufficiently large number  $M$  of diffraction patterns, the value of  $\langle A_n^2 \rangle_M$  approaches its statistical limit  $\langle A_n^2 \rangle$  [see Eqs. (11) and (13)]. This result immediately explains the advantage of calculation of the average CCFs in dilute systems as soon as it converges to a finite not fluctuating value. This opens the opportunity to determine the structural term  $\mathcal{L}^n(q)$  for systems with orientational disorder. At the same time our analysis shows that the favorable situation for such averaged analysis are dilute systems with a small number of orientations. If the number  $N$  of orientations increases, contribution of this averaged statistical term  $\langle A_n^2 \rangle_M$  decreases [for example, as  $\sim 1/N$  for a uniform distribution of orientations according to Eq. (13)], and the structural contribution  $\langle C_{q(clust)}^n \rangle_M$  can become smaller than the inter-cluster contribution  $\langle C_{q(int-clust)}^n \rangle_M$ . In Appendix B we give a detailed derivation for the statistical estimate of the inter-cluster contribution to  $\langle C_q^n \rangle_M$  [Eq. (20)].

One important question that is relevant to the discussion in this chapter is how many diffraction patterns  $M$  will be sufficient for calculation of the average  $\langle A_n^2 \rangle_M$  in order to reach reliably its statistical limit  $\langle A_n^2 \rangle$ . In other words, what is an averaging accuracy  $\varepsilon$  for a number  $M$  of diffraction patterns. This question can be addressed by establishing a confidence interval for  $\langle A_n^2 \rangle_M$ . Our analysis has shown (see Appendix C for details) that in the case of a uniform distribution of cluster orientations and averaging accuracy of 10% the probability that  $\langle A^2 \rangle_M$  lies in the interval from  $0.9\langle A^2 \rangle$  to  $1.1\langle A^2 \rangle$  is more than 90% for  $M = 10^3$ , and more than 99% for  $M = 10^4$  diffraction patterns.

We demonstrate these results by analyzing the ensemble averaged CCFs for three systems consisting of (a) 11, (b) 60 and (c) 121 clusters with the relative distances (a)  $\langle R \rangle/d \sim 5.0$ , (b)  $\langle R \rangle/d \sim 2.2$  and (c)  $\langle R \rangle/d \approx 1.5$  respectively. For each realization of a system in the ensemble different set of angles and cluster positions were considered. In Fig. 7 the results of the Fourier analysis of the CCFs averaged over  $M = 1000$  diffraction patterns are presented. According to our previous analysis, for this number of diffraction patterns the term  $\langle C_{q(clust)}^n \rangle_M$  in Eq. (20) with the probability more then 90% lies in the interval  $\langle C_{q(clust)}^n \rangle \pm 10\%$ .

Comparison of the averaged Fourier spectra for the system containing 11 clusters [Fig. 7(a)], with the ones calculated for different single realizations of the same system [Fig. 6] shows a strong enhancement of the contrast of the contribution from the internal structure of the pentagonal cluster, as compared to the inter-cluster contribution. At the same time, comparison of the magnitudes of the same Fourier components with the ones

obtained in a dilute limit for a system with the uniform distribution of cluster orientations [see Eq. (22) with  $M \rightarrow \infty$ ] reveals significant deviations [see Table III]. We attribute this effect to the presence of the inter-cluster contribution in the latter case.

Comparison of the results presented in Fig. 7 for three different samples shows, that as soon as the number of clusters in a dense system increases, the Fourier components  $\langle C_{q(clust)}^n \rangle_M$  related to the local structure decrease following the  $1/N$  dependence [see Eq. (13)]. It is well seen in Figs. 7(b) and 7(c) that for a system with the number of clusters  $N = 60$  the Fourier components with  $n = 10$  and 20 are already hardly resolved over the inter-cluster contribution, and for the system with  $N = 121$  clusters their contribution is not resolved over the level of inter-cluster contribution. Therefore, if the number of possible orientations of LSs in a dense system is sufficiently large, as in Fig. 7(c), calculation of the average CCFs may not provide information on the local structure in a disordered system.

The contribution of the inter-cluster correlations  $\langle C_{q(int-clust)}^n \rangle_M$  to the averaged CCF is well seen in an entire Fourier spectrum for all three systems presented in Fig. 7. In Appendix B an asymptotic estimate of the inter-cluster contribution in  $\langle C_q^n \rangle_M$  [Eq. (20)] is obtained (see solid red line in Fig. 7). As one can see in Fig. 7, our estimate of the inter-cluster contribution quite accurately reproduces the results of the direct calculations for all systems considered here. Our analysis has shown that for a sufficiently large number  $M$  of diffraction patterns the magnitude of the averaged inter-cluster contribution  $\langle C_{q(int-clust)}^n \rangle_M$  depends on the shape and size of the sample, as well as on the size of particles, and rather weakly on their number  $N$ .

The evolution of different terms in the expansion of the averaged Fourier component  $\langle C_q^n \rangle_M$  [Eq. (19)] as a function of the number  $M$  of diffraction patterns is presented in Fig. 9. The results are shown for  $C_q^n$  with  $n = 10$  for the same systems and  $q$  values as in Fig. 7. These results demonstrate the convergence of different terms to their average values. In particular, as one can see from Fig. 9, the averaged term  $\langle (S_2^n + S_3^n) / |I^0(q)|^2 \rangle_M$  in Eq. (19) asymptotically reaches zero and can be neglected after averaging over a few hundred diffraction patterns. For a system with the number of particles  $N = 11$  the structural term strongly dominates over the inter-cluster contribution at most of the  $q$  values (it is only lower for  $q = q_3$ ). Contrary to that case, for the systems with the number of clusters  $N = 60$  and  $N = 121$  the inter-cluster contribution term is larger than the term corresponding to



the structure of individual clusters for most of the  $q$ -values. That is the reason why the structural contribution could be hardly resolved for a system with  $N = 121$  clusters in Fig. 7(c).

Our results clearly show that, in the general case, the ensemble averaged CCF  $\langle C_q(\Delta) \rangle_M$  contains both the contribution from the individual clusters and the inter-cluster contribution [see Eq. (20)]. Therefore, the ensemble averaged CCF  $\langle C_q(\Delta) \rangle_M$  determined from a system of coherently illuminated clusters, in general, can not be considered as a function of only ‘single-particle quantities’ as it was stated in Refs.<sup>8,32</sup>. For a large number  $N$  of cluster orientations in the system the inter-cluster contribution  $\langle C_{q(int-clust)}^n \rangle_M$  can become dominant over the structural contribution  $\langle C_{q(clust)}^n \rangle_M$ . In the next section we consider a partially coherent scattering from disordered systems as one of the possible ways to suppress the inter-particle contribution to the CCF.

## VII. EFFECTS OF PARTIAL COHERENCE

In the previous sections we analyzed diffraction data obtained under conditions of coherent illumination of a disordered 2D sample. The influence of partial coherence of x-rays on the results of the x-ray cross-correlation analysis was briefly discussed in Ref.<sup>20</sup>. Here we will consider this question in more detail.

Partial coherence of x-rays can be characterized in terms of the mutual intensity function<sup>22,33</sup>  $J_{in}(\mathbf{r}_1, \mathbf{r}_2)$  of the beam incoming on the sample and defined as  $J_{in}(\mathbf{r}_1, \mathbf{r}_2) = \langle E(\mathbf{r}_1, t)E(\mathbf{r}_2, t) \rangle_T$ , where averaging is performed over times  $T$  much longer than the fluctuation time of the incoming x-ray field. It describes the statistical properties of the wavefield as a correlation function between two values of the electric field,  $E(\mathbf{r}_1, t)$  and  $E(\mathbf{r}_2, t)$ , at different points  $\mathbf{r}_1$  and  $\mathbf{r}_2$  in space and at the same time  $t$ , averaged over fluctuations of the wavefield. It is convenient also to introduce a normalized complex coherence factor,

$$\mu_{in}(\mathbf{r}_1, \mathbf{r}_2) = J_{in}(\mathbf{r}_1, \mathbf{r}_2) / [I_{in}(\mathbf{r}_1)I_{in}(\mathbf{r}_2)]^{1/2}, \quad (24)$$

where  $I_{in}(\mathbf{r}_1)$  and  $I_{in}(\mathbf{r}_2)$  are the intensity values of the incoming beam at points  $\mathbf{r}_1$  and  $\mathbf{r}_2$ , respectively, averaged over fluctuations of the wavefield. Using the definition (24), the effect of partial coherence of x-rays on the distribution of the scattered intensity  $I_{pcoh}(\mathbf{q})$  can be expressed in the far-field limit<sup>34</sup> as a convolution of the coherently scattered intensity  $I_{coh}(\mathbf{q})$

with the Fourier transform  $\mu_{\text{in}}(\mathbf{q})$  of the complex coherence factor<sup>35</sup>,

$$I_{\text{pcoh}}(\mathbf{q}) = I_{\text{coh}}(\mathbf{q}) \otimes \mu_{\text{in}}(\mathbf{q}). \quad (25)$$

One can directly use this equation for calculations of the distribution of the scattered intensity for partially coherent illumination of a sample, as soon as coherence properties of the incoming beam are defined by the known complex coherence factor.

We will demonstrate the influence of partial coherence of x-rays on the results of the XCCA by analyzing the scattering experiment with different degree of spatial coherence of the incident x-rays for a disordered system consisting of 11 clusters, presented in Fig. 6(c) (for Sample 3). We assume in our simulations a Gaussian form of the complex coherence factor (so-called Gaussian Shell-model<sup>33</sup>)

$$\mu_{\text{in}}(\mathbf{r}_1, \mathbf{r}_2) = \exp[-(\mathbf{r}_1 - \mathbf{r}_2)^2 / 2l_{\text{coh}}^2], \quad (26)$$

where  $l_{\text{coh}}$  is the transverse coherence length. Three different values of the transverse coherence length,  $l_{\text{coh}} = 1.2 \mu\text{m}$ ,  $600 \text{ nm}$ , and  $300 \text{ nm}$  are considered in our calculations. These values of the coherence length  $l_{\text{coh}}$  allow to probe typical length scales in a chosen sample, starting from comparably large coherent length, going down to the size of a single pentagonal cluster and below. In Fig. 8 diffraction patterns calculated according to Eq. (25,26) as a function of the transverse coherence length  $l_{\text{coh}}$ , and corresponding Fourier spectra of the CCFs calculated at three different  $q$  values for each pattern are presented. As one can see from Fig. 8, decrease of the transverse coherence length corresponds to smearing of speckles in a diffraction pattern. At the same time, smearing of speckles reduces the contribution from inter-cluster correlations in the Fourier spectra of the CCFs. Partial coherence of the incident x-rays acts as a filter for the Fourier components of the CCF with a threshold defined by the value of the transverse coherence length. Decrease of the value of  $l_{\text{coh}}$  down to the size of a single pentagonal cluster considerably reduces the high-frequency contribution in the Fourier spectra, however, the contributions from the inter-cluster correlations do not vanish completely [compare Fig. 8(b) with the results for the case of fully coherent illumination of the Sample 3 in Fig. 6(c)]. Further decrease of the value of  $l_{\text{coh}}$  below the size of a single cluster leads to a further refinement of the Fourier spectra, and the contribution from the local structure becomes clearly dominant [Fig. 8(c)].

Comparison of the Fourier components of the CCFs  $C_q^n$  ( $n = 10, 20$ ) for the case of fully coherent scattering ( $C_q^{\text{coh}}$ ,  $l_{\text{coh}} = \infty$ ) and partially coherent scattering ( $C_q^{\text{pcoh}}$ ,  $l_{\text{coh}} =$

600 nm) from a system of 11 clusters with a uniform distribution of orientations is presented in Table IV. Note, that absolute values of the Fourier components  $C_q^n{}^{\text{pcoh}}$  corresponding to partially coherent data are significantly lower at all  $q$  values, as compared to  $C_q^n{}^{\text{coh}}$  values. At the same time, the inter-cluster contribution is suppressed significantly stronger and as a result one gets higher contrast for the LS contribution.

It is interesting to compare these results with the ones presented in Ref.<sup>35</sup>, where the effects of a partial coherence were studied on a model sample represented by an ordered array of identical particles in the same orientation. It was demonstrated<sup>35</sup>, that when the transverse coherence length of the incoming beam illuminating such a sample approaches the size of a single particle, the resulting diffraction pattern looks similar to the one produced by a single particle. Taking this into account, one would expect that for the disordered sample, considered here, a decrease of the transverse coherence length down to the size of a single pentagonal cluster will lead to a diffraction pattern, which resembles a sum of 11 diffraction patterns obtained for a single cluster in 11 different orientations. Such a diffraction pattern with the partial illumination of x-rays could be sufficient for the future phase retrieval analysis<sup>1,2,7,8</sup>.

Analysis of the effect of partial coherence on the results of the cross-correlation analysis allows the following conclusions to be made for disordered 2D systems. If the value of the transverse coherence length is of the order of a size of a single particle (cluster, molecule), the contribution of the inter-particle correlations strongly decreases (completely vanishing in the case of a dilute system), enabling easier identification of the Fourier components related to the internal structure of particles. Therefore, one may prefer to perform a scattering experiment with partially coherent x-ray beam, in order to extract the information on local structure of a disordered system. On the other hand, varying a degree of coherence of x-rays, one could probe the structural correlations in a disordered system at different length scales, which may be particularly useful in studies of a medium range order.

## VIII. CONCLUSIONS

In a summary, our simulations have demonstrated that in order to get reliable information about the local structures the systems with a high degree of orientational order are preferable ones comparing to the completely disordered systems. Due to the statistical nature of the

CCFs the information on local structure of a completely disordered system from a single diffraction pattern is practically not accessible. Averaging of CCFs over an ensemble of diffraction patterns allows to overcome this difficulty and gives reliable information on the structure of clusters forming the system. An estimate of a number of diffraction patterns necessary to determine an averaged CCF with a certain accuracy was obtained. Our analysis shows that few tens of thousands of diffraction patterns have to be averaged to get a reliable result.

The limiting factor for this analysis becomes the number of particles in the system and the inter-particle correlations that are always present in the conditions of coherent illumination. Our simulations show that the systems with a small number of clusters are preferable. Averaging of CCFs over an ensemble of diffraction patterns for such systems can significantly enhance the contrast of the structural contribution as compared to the inter-particle contribution. For the systems with a large number of clusters this averaging procedure could become not efficient due to the strongly suppressed signal (it scales as  $1/N$  with the number of particles  $N$ ) from the structural contribution. Our analysis has shown that the averaged inter-particle contribution practically does not depend on the number of particles and their density, but rather depends on the size and the shape of the system as well as on the size of clusters composing the system. Utilizing partially coherent incident beams with the coherence length about the size of the clusters composing the system could suppress the inter-particle contribution and make possible determination of the local structure of a disordered system.

Interesting open question left for the future work is an extension of XCCA to the three-dimensional (3D) systems. The additional degree of freedom for orientation of 3D clusters introduces a significant complication in the extraction of the structural information from the XCCA analysis. As it was pointed out in our previous work<sup>20</sup> one possible solution can be the measurements performed at high scattering angles when the curvature of the Ewald sphere has to be taken into account. This will be especially interesting case for atomic systems (organic and non-organic molecules, biological systems, etc.) as opposed to the systems composed of colloidal particles considered in this work. The full cross-correlation analysis for such practically important systems could open new horizons in exploring the nature of disordered systems.

## ACKNOWLEDGMENTS

We acknowledge a careful reading of the manuscript by H. Franz and A. Bell. Part of this work was supported by BMBF Proposal 05K10CHG Coherent Diffraction Imaging and Scattering of Ultrashort Coherent Pulses with Matter in the framework of the German-Russian collaboration Development and Use of Accelerator-Based Photon Sources.

---

\* Reference author: ivan.vartanians@desy.de

- <sup>1</sup> Z. Kam, *Macromolecules*, **10**, 927 (1977).
- <sup>2</sup> Z. Kam, *J. Theor. Biol.*, **82**, 15 (1980).
- <sup>3</sup> W. Ackermann *et al.*, *Nature Photonics*, **1**, 336 (2007).
- <sup>4</sup> P. Emma *et al.*, *Nature Photonics*, **4**, 641 (2010).
- <sup>5</sup> T. Tanaka and T. Shintake, *SCSS XFEL conceptual design report* (2005), <http://www-xfel.spring8.or.jp/SCSSCDR.pdf>.
- <sup>6</sup> M. Altarelli *et al.*, *The European X-Ray Free-Electron Laser Technical design report* (2007), [http://xfel.desy.de/technical\\_information/tdr/tdr/](http://xfel.desy.de/technical_information/tdr/tdr/).
- <sup>7</sup> D. K. Saldin, V. L. Shneerson, M. R. Howells, S. Marchesini, H. N. Chapman, M. Bogan, D. Shapiro, R. A. Kirian, U. Weierstall, K. E. Schmidt, and J. C. H. Spence, *New J. Phys.*, **12** (2010).
- <sup>8</sup> D. K. Saldin, H. C. Poon, V. L. Shneerson, M. Howells, H. N. Chapman, R. A. Kirian, K. E. Schmidt, and J. C. H. Spence, *Phys. Rev. B*, **81** (2010).
- <sup>9</sup> S. R. Elliott, *Nature*, **354**, 445 (1991).
- <sup>10</sup> H. W. Sheng, W. K. Luo, F. M. Alamgir, J. M. Bai, and E. Ma, *Nature*, **439**, 419 (2006).
- <sup>11</sup> M. M. J. Treacy, J. M. Gibson, L. Fan, D. J. Paterson, and I. McNulty, *Rep. Prog. Phys.*, **68**, 2899 (2005).
- <sup>12</sup> H. Shintani and H. Tanaka, *Nature Physics*, **2**, 200 (2006).
- <sup>13</sup> T. Kawasaki, T. Araki, and H. Tanaka, *Phys. Rev. Lett.*, **99** (2007).
- <sup>14</sup> H. Tanaka, T. Kawasaki, H. Shintani, and K. Watanabe, *Nature Materials*, **9**, 324 (2010).
- <sup>15</sup> A. S.-Y. Sheu and S. Rice, *J. Chem. Phys.*, **128**, 244517 (2008).
- <sup>16</sup> A. S.-Y. Sheu and S. Rice, *J. Chem. Phys.*, **129**, 124511 (2008).

- <sup>17</sup> R. Pindak, D. E. Moncton, S. C. Davey, and J. W. Goodby, Phys. Rev. Lett., **46**, 1135 (1981).
- <sup>18</sup> P. Wochner, C. Gutt, T. Autenrieth, T. Demmer, V. Bugaev, A. Diaz-Ortiz, A. Duri, F. Zontone, G. Grübel, and H. Dosch, Proc. Nat. Acad. Sci., **106**, 11511 (2009).
- <sup>19</sup> P. Wochner, M. Castro-Colin, S. Bogle, and V. Bugaev, Int. J. Mat. Res., **102**, 874 (2011).
- <sup>20</sup> M. Altarelli, R. P. Kurta, and I. A. Vartanians, Phys. Rev. B, **82** (2010).
- <sup>21</sup> In this paper we consider, for simplicity, the reduced form of the CCF  $C_{q_1, q_2}(\Delta)$  introduced in Ref.<sup>20</sup> with  $q_1 = q_2 = q$ . Note, that in some cases, for example in Ref.<sup>8</sup>, a not normalized form of the CCF  $C_{q_1, q_2}(\Delta)$  is used. In this case the influence of the form-factors of the individual scatterers on the form of CCFs is not excluded by normalization.
- <sup>22</sup> J. Goodman, *Statistical Optics* (Wiley, New York, 2000).
- <sup>23</sup> J. Goodman, *Speckle Phenomena in Optics: theory and applications* (Roberts and Company Publishers, Englewood, Colorado, 2007).
- <sup>24</sup> Note, that in Eq. (14) we combined the complex-valued terms 2 and 3 from Ref.<sup>20</sup> into one real-valued term.
- <sup>25</sup> All simulations of diffraction patterns were performed using the computer code MOLTRANS.
- <sup>26</sup> Sufficient resolution of the detector is important for a reliable determination of all contributions to the CCF defined by Eqs. (14, 15a-15c).
- <sup>27</sup> H. Reichert, O. Klein, H. Dosch, M. Denk, V. Honkimäki, T. Lippmann, and G. Reiter, Nature, **408**, 839 (2000).
- <sup>28</sup> A. Okabe, B. Boots, K. Sugihara, and S. N. Chiu, *Spatial tessellations: concepts and applications of Voronoi diagram*, Wiley series in probability and statistics (John Wiley and Sons Ltd., Chichester, West Sussex, 2000).
- <sup>29</sup> For the given sample area ( $10 \times 10 \mu\text{m}^2$ ) and cluster size ( $d = 600 \text{ nm}$ ) one can simulate a close-packed system composed of more than 200 clusters. However, such a high density of clusters is not convenient for modeling of a disordered system since it may induce some positional ordering in the system. To achieve better positional disorder and at the same time to keep a relatively high density of clusters we consider here the systems composed of 121 clusters.
- <sup>30</sup> Since a pentagonal cluster has 5-fold rotational symmetry, any rotation of the cluster about its 5-fold rotational axis can be associated with the angular orientation in the interval  $(-\pi/5, \pi/5)$ . In other words, any rotation  $-\infty < \phi_k < \infty$  can be expressed as an orientation  $\phi_k^0 = \phi_k \bmod \pi/5$ , which is defined in the interval  $-\pi/5 \leq \phi_k^0 < \pi/5$ . That is why, it is sufficient for a pen-

tagonal cluster to consider the uniform distribution of its angular orientations in the interval  $(-\pi/5, \pi/5)$ . However, one should be careful in simulations where clusters have a nonuniform distribution of orientations. For example, in the case of a broad Gaussian distribution of orientations of pentagonal clusters, where a possible rotation of the cluster may fall outside the interval  $-\pi/5 \leq \phi_k < \pi/5$ , one should rather deal with a wrapped distribution of orientations<sup>36,37</sup>, which falls into the scope of directional statistics. In this paper we choose only such sets of orientations of pentagonal clusters (both for the uniform and for the Gaussian distributions), where all the angles satisfy the condition  $-\pi/5 \leq \phi_k < \pi/5$ .

- <sup>31</sup> D. K. Saldin, H. C. Poon, M. J. Bogan, S. Marchesini, D. A. Shapiro, R. A. Kirian, U. Weierstall, and J. C. H. Spence, *Phys. Rev. Lett.*, **106** (2011).
- <sup>32</sup> H. C. Poon and D. K. Saldin, *Ultramicroscopy*, **111**, 798 (2011).
- <sup>33</sup> L. Mandel and E. Wolf, *Optical Coherence and Quantum Optics* (Cambridge University Press, Cambridge, 1995).
- <sup>34</sup> Here, we considered a broad distribution of the intensity  $I_{\text{in}}(\mathbf{r}_1) = I_{\text{in}}(\mathbf{r}_2) = \text{const}$  over the size of the sample and a spatially uniform distribution of the complex coherence factor  $\mu_{\text{in}}(\mathbf{r}_1, \mathbf{r}_2) = \mu_{\text{in}}(\mathbf{r}_2 - \mathbf{r}_1)$  (Shell-model).
- <sup>35</sup> I. Vartanyants and I. Robinson, *J. Synchrotron Rad.*, **10**, 409 (2003).
- <sup>36</sup> K. V. Mardia and P. E. Jupp, *Directional Statistics*, Wiley series in probability and statistics (John Wiley and Sons Ltd., Chichester, West Sussex, 2000).
- <sup>37</sup> S. R. Jammalamadaka and A. Sengupta, *Topics in circular statistics*, Series on multivariate analysis, Vol. 5 (World Scientific Publishing Company, River Edge, N.J., 2001).
- <sup>38</sup> Due to normalization the scattering factors cancel out in the expression for  $C_{q(int-clust)}^n$  in Eq. (B1) for identical particles forming clusters.
- <sup>39</sup> D. Rasch, *Einführung in die mathematische Statistik* (VEB Deutscher Verlag der wissenschaften, Berlin, 1989).
- <sup>40</sup> K. Kodama, S. Iikubo, T. Taguchi, and S. Shamoto, *Acta Cryst. A*, **62**, 444 (2006).
- <sup>41</sup> A. Papoulis, *Probability, random variables, and stochastic processes* (McGraw-Hill, Inc., New York, 1991).

## Appendix A:

Here we present the results of calculations of the function  $L^n(q)$  [Eq. (7)] for a centered pentagonal 2D cluster [Fig. 1(b)]. The direction of the incident x-ray beam is assumed to be parallel to 5-fold rotational axis of a cluster. We define the electron density of a cluster as a real-valued quantity in the following form

$$\tilde{\rho}(\mathbf{r}) = \sum_{i=1}^{N_s} f_i(q) \delta(\mathbf{r} - \mathbf{r}_i), \quad (\text{A1})$$

where  $f_i(q)$  and  $\mathbf{r}_i$  are the form factor and the radius vector of the  $i$ -th scatterer and  $N_s$  is the number of scatterers in the cluster. Substituting Eq. (A1) into Eq. (7) and assuming that the cluster is composed of scatterers with the same form factor  $f(q)$  we obtain

$$L^n(q) = |f(q)|^2 \sum_{s,t} J_n(q|\mathbf{r}_{st}|) e^{-in\phi_{\mathbf{r}_{st}}} \quad (\text{A2})$$

Using the coordinates of scatterers in Eq. (A2) we obtain for a centered pentagonal 2D cluster

$$L^n(q) = \begin{cases} |f(q)|^2 \{6J_n(0) + 10[J_n(qa) + J_n(A_1qa) + J_n(A_2qa)]\} & \text{if } n = 0, \\ & n \bmod 20 = 0; \\ 10|f(q)|^2 [J_n(qa) - J_n(A_1qa) - J_n(A_2qa)] & \text{if } n \bmod 10 = 0, \\ & n \bmod 20 \neq 0; \\ 0 & \text{other } n, \end{cases} \quad (\text{A3})$$

where  $a$  is a distance between centers of the nearest particles inside the pentagonal cluster,  $A_1 = \sqrt{\frac{1}{2}(5 - \sqrt{5})} \approx 1.18$ , and  $A_2 = \sqrt{\frac{1}{2}(5 + \sqrt{5})} \approx 1.9$ . Note, that the function  $L^n(q)$  presented here for a *centered* pentagonal cluster differs from the one defined in Ref.<sup>20</sup> for a LS with a fivefold symmetry due to the missing central particle in the latter case.

In our simulations we consider the following form factor of a spherical particle

$$f(q) = \rho \frac{4\pi [\sin(qr) - (qr) \cos(qr)]}{q^3}, \quad (\text{A4})$$

where  $\rho$  is the average electron density inside a particle and  $r$  is the radius of a particle.

## Appendix B:

Here we determine a statistical estimate of the average Fourier components of the inter-cluster contribution  $\langle C_{q(int-clust)}^n \rangle$  [see Eq. (20)]. First, we will consider a simplified model of



a disordered system, which consist of spherical particles. Next, we will show by simulations that asymptotic equations derived for this simplified model can be used to determine the inter-particle contribution in a disordered system composed of more complicated objects.

According to Eq. (21b) the normalized Fourier component  $C_{q(int-clust)}^n$  can be written as follows

$$C_{q(int-clust)}^n = S_4^n / [I^0(q)]^2, \quad (B1)$$

where

$$S_4^n = 4 \left| \sum_{k_2 > k_1} L_{k_1, k_2}^n(q) \right|^2, \quad (B2a)$$

$$|I^0(q)|^2 = \left| \sum_{k_1, k_2} L_{k_1, k_2}^0(q) \right|^2 = \left| N L^0(q) + 2 \sum_{k_2 > k_1} L_{k_1, k_2}^0(q) \right|^2. \quad (B2b)$$

Here  $I^0(q)$  was decomposed into contribution of two terms with  $k_1 = k_2$  and  $k_1 \neq k_2$  similar to the expression in Eq. (4),  $L_{k_1, k_2}^n(q)$  is defined by Eq. (5b), and  $L^0(q)$  by Eq. (7) with  $n = 0$ .

Using the expression for the electron density  $\tilde{\rho}(\mathbf{r})$  [see Eq. (A1)] in Eqs. (5b) and (7) for a system of clusters composed of identical scatterers we can write<sup>20</sup>

$$L_{k_1, k_2}^n(q) = \sum_{s, t} J_n(q | \mathbf{R}_{k_2, k_1}^{s, t} |) e^{-in\phi_{\mathbf{R}_{k_2, k_1}^{s, t}}}, \quad (B3a)$$

$$L^0(q) = \sum_{s, t} J_0(q | \mathbf{r}_{st} |). \quad (B3b)$$

In Eq. (B3a) the summation is performed over positions of the scatterers inside the clusters  $k_1$  and  $k_2$ , and in Eq. (B3b) inside a single cluster<sup>38</sup>.

Now, we consider a simplified model of a disordered system, where we substitute a cluster by a spherical particle of the same size as the original cluster. In this case  $\mathbf{r}_{st} = 0$  and Eqs. (B3a, B3b) reduce to

$$L_{k_1, k_2}^n(q) = J_n(q | \mathbf{R}_{k_2, k_1} |) e^{-in\phi_{\mathbf{R}_{k_2, k_1}}}, \quad (B4a)$$

$$L^0(q) = 1. \quad (B4b)$$

Further, the double summation in  $S_4^n$  and  $I^0(q)$  [Eqs. (B2a) and (B2b)] can be substituted by a single sum over the same set of vectors. Preserving the number of contributions in the

original sum we can write

$$S_4^n = 4 \left| \sum_{k=1}^{N_1} J_n(q|\mathbf{R}_k|) e^{-in\phi_{\mathbf{R}_k}} \right|^2, \quad (\text{B5a})$$

$$[I^0(q)]^2 = \left[ N + 2 \sum_{k=1}^{N_1} J_0(q|\mathbf{R}_k|) \right]^2, \quad (\text{B5b})$$

where  $N_1 = N(N-1)/2$ .

The sum in Eq. (B5a) is, in fact, a random phasor sum<sup>22,23</sup> defined as

$$\mathbf{B}_n = B_n e^{i\theta_n} = \frac{1}{N_1} \sum_{k=1}^{N_1} b_k^n e^{in\phi_k}, \quad n \neq 0, \quad (\text{B6})$$

where  $B_n$  and  $\theta_n$  are the length and phase of the random phasor sum  $\mathbf{B}_n$ , and the amplitude and phase of each term are given by

$$b_k^n \equiv J_n(q|\mathbf{R}_k|), \quad \phi_k \equiv \phi_{\mathbf{R}_k}. \quad (\text{B7})$$

With these definitions we can also formally write the random sum in Eq. (B5b) in the following form,

$$B_0 = \frac{1}{N_1} \sum_{k=1}^{N_1} b_k^0. \quad (\text{B8})$$

Substituting Eqs. (B6 - B8) into Eqs. (B5a, B5b) we obtain

$$S_4^n = 4N_1^2 B_n^2, \quad (\text{B9a})$$

$$[I^0(q)]^2 = [N + 2N_1 B_0]^2 = N^2 + 4NN_1 B_0 + 4N_1^2 B_0^2. \quad (\text{B9b})$$

As soon as we are interested in averaged characteristics of the Fourier component  $C_{q(int-clust)}^n$  [Eq. (B1)] we note that both terms  $S_4^n$  and  $[I^0(q)]^2$  can be considered as random variables as well. Applying the theorem on a linearized approximation<sup>39</sup> to evaluate  $\langle C_{q(int-clust)}^n \rangle$  we can express an average of the ratio of these two variables as a ratio of the corresponding averages

$$\langle C_{q(int-clust)}^n \rangle \approx \langle S_4^n \rangle / \langle [I^0(q)]^2 \rangle, \quad (\text{B10})$$

where

$$\langle S_4^n \rangle = 4N_1^2 \langle B_n^2 \rangle, \quad (\text{B11a})$$

$$\langle [I^0(q)]^2 \rangle = N^2 + 4NN_1 \langle B_0 \rangle + 4N_1^2 \langle B_0^2 \rangle. \quad (\text{B11b})$$

It is clear that this approach is valid for  $q$ -values when  $\langle [I^0(q)]^2 \rangle \neq 0$ . Our direct simulations have shown that approximation described by Eq. (B10) is well satisfied for our systems.

According to Eqs. (B10), (B11a) and (B11b), in order to estimate the value  $\langle C_{q(int-clust)}^m \rangle$  we need to determine the statistical averages  $\langle B_n^2 \rangle$  and  $\langle B_0 \rangle$ . Following the approach of Ref.<sup>23</sup>, one can obtain the marginal probability density function for the random variable  $B_n^2$  ( $n \neq 0$ ), that reduces for a large number of particles to a normal distribution

$$P(B_n^2) = \frac{1}{2\sigma_n^2} \exp\left(-\frac{B_n^2}{2\sigma_n^2}\right), \quad B_n^2 \geq 0, \quad (\text{B12})$$

where

$$\sigma_n^2 = \frac{1}{N_1^2} \sum_{k=1}^{N_1} \frac{\langle (b_k^n)^2 \rangle}{2} = \frac{\langle J_n^2(q|\mathbf{R}|) \rangle}{2N_1}. \quad (\text{B13})$$

Using the probability density function  $P(B_n^2)$  the mean value of the random variable  $B_n^2$  and its variance can be determined for  $n \neq 0$  as

$$\langle B_n^2 \rangle = 2\sigma_n^2 = \langle J_n^2(q|\mathbf{R}|) \rangle / N_1, \quad (\text{B14a})$$

$$\sigma_{B_n^2}^2 = 4\sigma_n^4 = [\langle J_n^2(q|\mathbf{R}|) \rangle / N_1]^2. \quad (\text{B14b})$$

According to Eqs. (B14a) and (B14b) the value  $B_n^2$  fluctuates around the mean value  $\langle B_n^2 \rangle$  with the standard deviation  $\sigma_{B_n^2}$ . According to Eq. (B11a), this also determines the statistical behaviour of the term  $S_4^n$ . Note, that similarly to the random phasor sum considered for LS contribution, averaging over sufficiently large number  $M$  of diffraction patterns leads to a decrease of the variance  $\sigma_{B_n^2}^2$ , and the result of such averaging  $\langle B_n^2 \rangle_M$  asymptotically approaches its statistical limit, i.e.  $\langle B_n^2 \rangle_M \rightarrow \langle B_n^2 \rangle$  (see Appendix C).

The average values  $\langle B_0 \rangle$  and  $\langle B_0^2 \rangle$  can be directly determined as

$$\langle B_0 \rangle = \frac{1}{N_1} \sum_{k=1}^{N_1} \langle b_k^0 \rangle = \langle J_0(q|\mathbf{R}|) \rangle, \quad (\text{B15a})$$

$$\begin{aligned} \langle B_0^2 \rangle &= \frac{1}{N_1^2} \sum_{k_1, k_2} \langle b_{k_1}^0 b_{k_2}^0 \rangle = \frac{1}{N_1^2} \left[ \sum_{k=1}^{N_1} \langle J_0^2(q|\mathbf{R}_k|) \rangle + \sum_{k_1 \neq k_2} \langle J_0(q|\mathbf{R}_{k_1}|) J_0(q|\mathbf{R}_{k_2}|) \rangle \right] \\ &= \frac{1}{N_1} [\langle J_0^2(q|\mathbf{R}|) \rangle + (N_1 - 1) \langle J_0(q|\mathbf{R}|) \rangle^2], \end{aligned} \quad (\text{B15b})$$

In the derivation of Eq. (B15b) we assumed that for  $k_1 \neq k_2$ ,  $\langle J_0(q|\mathbf{R}_{k_1}|) J_0(q|\mathbf{R}_{k_2}|) \rangle = \langle J_0(q|\mathbf{R}_{k_1}|) \rangle \langle J_0(q|\mathbf{R}_{k_2}|) \rangle$ .

Using these results in Eqs. (B11a) and (B11b) we obtain for the statistical averages  $\langle S_4^n \rangle$  and  $\langle [I^0(q)]^2 \rangle$ ,

$$\langle S_4^n \rangle = 4N_1 \langle J_n^2(q|\mathbf{R}|) \rangle, \quad (\text{B16a})$$

$$\langle [I^0(q)]^2 \rangle = N^2 + 4N_1 [N \langle J_0(q|\mathbf{R}|) \rangle + \langle J_0^2(q|\mathbf{R}|) \rangle + (N_1 - 1) \langle J_0(q|\mathbf{R}|) \rangle^2]. \quad (\text{B16b})$$

For a large number of particles in the system  $N \gg 1$ ,  $N_1 \approx N^2/2$  and substituting now Eqs. (B16a, B16b) into Eq. (B10) we obtain for the statistical average

$$\langle C_{q \text{ (int-clust)}}^n \rangle = \frac{2 \langle J_n^2(q|\mathbf{R}|) \rangle}{1 + 2N \langle J_0(q|\mathbf{R}|) \rangle + 2 \langle J_0^2(q|\mathbf{R}|) \rangle + N^2 \langle J_0(q|\mathbf{R}|) \rangle^2}. \quad (\text{B17})$$

Our direct simulations for the systems with a large number of clusters considered here show, that the terms  $N \langle J_0(q|\mathbf{R}|) \rangle$ ,  $\langle J_0^2(q|\mathbf{R}|) \rangle$  as well as  $N^2 \langle J_0(q|\mathbf{R}|) \rangle^2$  in the denominator of Eq. (B17) are much smaller than unity and Eq. (B17) reduces to the following asymptotic expression

$$\langle C_{q \text{ (int-clust)}}^n \rangle \approx 2 \langle J_n^2(q|\mathbf{R}|) \rangle. \quad (\text{B18})$$

As we can see from this expression, for the systems considered here the inter-cluster contribution practically does not depend on the number of clusters  $N$ . Substituting Eqs. (B18) and (22) in Eq. (20) for a sufficiently large number of diffraction pattern we obtain

$$\langle C_q^n \rangle = \mathcal{L}^n(q) \cdot \langle A_n^2 \rangle + 2 \langle J_n^2(q|\mathbf{R}|) \rangle. \quad (\text{B19})$$

Taking into account that for a uniform distribution of orientations  $\langle A_n^2 \rangle = 1/N$  [Eq. (13)] we obtain the upper limit for the number of particles  $N$ , for which the structural contribution dominates over the inter-cluster contribution. The first term in Eq. (B19) strongly dominates, if the number of clusters  $N$  satisfies the following condition

$$N \ll \frac{\mathcal{L}^n(q)}{2 \langle J_n^2(q|\mathbf{R}|) \rangle}. \quad (\text{B20})$$

For the systems considered here this estimate shows that the number of particles in the system should be considerably lower than about a few hundreds, that is in a good agreement with our observations.

The averages  $\langle J_0(q|\mathbf{R}|) \rangle$  and  $\langle J_n^2(q|\mathbf{R}|) \rangle$  (for all  $n$  values) required to calculate  $\langle C_{q \text{ (int-clust)}}^n \rangle$  can be determined using the radial distribution function (RDF)  $P(R)$

$$\langle J_0(q|\mathbf{R}|) \rangle = \int_{R_{\min}}^{R_{\max}} P(R) J_0(qR) dR, \quad \langle J_n^2(q|\mathbf{R}|) \rangle = \int_{R_{\min}}^{R_{\max}} P(R) J_n^2(qR) dR, \quad (\text{B21})$$

where  $R_{\min}$  and  $R_{\max}$  are the minimum and maximum inter-particle distances in the system. According to its definition<sup>40</sup>,  $P(R)$  gives a number of particles in an annulus of thickness  $dR$  at a distance  $R$  from another particle. Note, that  $P(R)$  defined in such a way should be normalized by the number of particles  $N$  before using it in Eq. (B21), in order to satisfy the condition  $\int_{R_{\min}}^{R_{\max}} P(R)dR = 1$ . In our simulations we consider a statistical distribution of spherical particles with a diameter  $d$  within the square 2D sample with the length  $l$  of a side of the square. For such a system, it is convenient to determine the RDF using the “partial radial distribution function”<sup>40</sup> (PRDF)  $P_{\text{part}}(R, u, v)$ ,

$$P(R) = \frac{1}{S_{fs}} \int_{u=d/2}^{l-d/2} \int_{v=d/2}^{l-d/2} P_{\text{part}}(R, u, v) du dv, \quad d \leq R \leq \sqrt{2}(l - d), \quad (\text{B22})$$

where  $P_{\text{part}}(R, u, v)$  defines the number of particles in an annulus of thickness  $dR$  at a distance  $R$  from a particle at a position  $P$  which is located from the sides of the square at distances  $u$  and  $v$  [see Fig. 10], and  $S_{fs} = (l - d)^2 - \pi d^2$  is a part of the sample area accessible for  $P_{\text{part}}(R, u, v)$ . Depending on the radius  $R$  and the position  $P$ ,  $P_{\text{part}}(R, u, v)$  can be proportional to the length of the circumference [Fig. 10(a)], the length of a single arc [Fig. 10(b)], or the total length of the few arcs [Fig. 10(c)]. Clearly, the RDF depends on the size and the shape of the particles and the sample. Eq. (B22) ensures that the minimum distance between the particles is  $d$ , and minimum distance from a particle center to the sample edge is  $d/2$ ; it also defines the region of integration in Eq. (B21) with  $R_{\min} = d$ ,  $R_{\max} = \sqrt{2}(l - d)$ . The RDF  $P(R)$  calculated using Eq. (B22) for a system of spherical particles with  $d = 600$  nm and the size of the sample  $l = 10$   $\mu\text{m}$  is presented in Fig. 10(d). As we can see from this figure, the maximum of this distribution lies between 4  $\mu\text{m}$  and 6  $\mu\text{m}$  that corresponds approximately to the half size of the system.

We compare now the results of an asymptotic estimate  $\langle C_{q \text{ (int-clust)}}^n \rangle$  of the inter-cluster contribution for a system composed of spherical particles using Eqs. (B10), (B16a) and (B16b) with the results of direct calculations for different systems by averaging  $C_{q \text{ (int-clust)}}^n$  in Eq. (B1) over  $M = 1000$  realizations. In the latter case, the coordinates of particles in each realization were used in calculations of Eqs. (B2a) and (B2b). Three 2D disordered systems composed of (a) spherical particles [Fig. 10(e)], (b) triangular clusters [Fig. 10(f)] and (c) centered pentagonal clusters [Fig. 10(g)] were considered in these direct calculations.

Fourier components  $\langle C_{q \text{ (int-clust)}}^n \rangle_M$  and  $\langle C_{q \text{ (int-clust)}}^n \rangle$  calculated using Eqs. (B1) and (B10) for  $n = 4$  and different  $q$  values for systems with  $N = 11, 60$  and 121 particles are

presented in Fig. 11. The Fourier component with  $n = 4$  was considered because for the chosen symmetry of clusters only the term  $S_4^n$  in Eq. (14) contributes to  $C_q^n$  [terms  $S_1^n$  and  $(S_2^n + S_3^n)$  are equal to zero]. In all three cases direct calculations show that the inter-cluster contribution practically does not depend on the internal structure of particles. As one can see from Fig. 11, the inter-cluster contribution can be accurately estimated using the asymptotic expression. The inter-cluster contribution calculated using the asymptotic expression [Eqs. (B10), (B16a) and (B16b)] for three systems considered here for an entire Fourier spectrum is presented in Fig. 7 with the solid red line. As one can see, our estimate quite accurately reproduces the results of direct calculations for all systems considered here in a wide range of  $n$  values. We want to note here, that asymptotic values for the system with  $N = 11$  clusters slightly overestimate contribution from the inter-cluster correlation [see Fig. 7(a) and Fig. 11(a)] that is due to the violation of the central limit theorem, used in the derivation of Eq. (B10).

Finally, our analysis shows that the asymptotic expression for the average inter-particle contribution  $\langle C_{q(int-clust)}^n \rangle$  [Eqs. (B10), (B16a) and (B16b)] can be effectively used as an estimate of this contribution to the total Fourier component  $\langle C_q^n \rangle_M$  in Eq. (20). We demonstrated that  $\langle C_{q(int-clust)}^n \rangle$  depends on the shape and size of the sample, as well as the size of particles, and rather weakly on their number  $N$  in the system.

### Appendix C:

Here, we give an estimate of the number of diffraction patterns  $M$  that is required to determine the ensemble averaged amplitude  $\langle A_n^2 \rangle_M$  [Eq. (23)] with a given accuracy  $\varepsilon$ . We define  $\varepsilon$  as a maximum deviation of  $\langle A_n^2 \rangle_M$  from its statistical limit  $\langle A_n^2 \rangle$ .

Each single x-ray pulse produces a diffraction pattern from a static distribution of clusters in the system, and for the  $m$ -th diffraction pattern the square modulus  $A_n^2(m)$  of the random phasor sum (10) can be written as

$$A_n^2(m) = \langle A_n^2 \rangle + \delta A_n^2(m), \quad (\text{C1})$$

where  $\langle A_n^2 \rangle$  is the statistical average and  $\delta A_n^2(m)$  describes fluctuations of this average value from one diffraction pattern to another.

Introducing an average over  $M$  diffraction patterns one can write

$$\langle A^2(m) \rangle_M = \langle A^2 \rangle + \langle \delta A^2(m) \rangle_M. \quad (\text{C2})$$

Hereafter we omit the subscript  $n$  for brevity. The average  $\langle A^2(m) \rangle_M$  can be considered as a random variable with a variance  $\sigma_M^2$ . We assume, that the measurement of  $A^2(m)$  is a stationary ergodic process, and  $\sigma_{M \rightarrow \infty}^2 \rightarrow 0$ . For a sufficiently large but finite number of diffraction patterns  $M$  satisfying ergodicity condition, the variance  $\sigma_M^2$  can be estimated as<sup>41</sup>

$$\sigma_M^2 \simeq \frac{1}{M} \sum_{\Delta m=0}^M C(\Delta m). \quad (\text{C3})$$

Here  $C(\Delta m)$  ( $\Delta m = m - m'$ ) is the autocovariance of  $A^2(m)$ ,

$$C(\Delta m) = \langle A^2(m) A^2(m + \Delta m) \rangle - \langle A^2 \rangle^2, \quad (\text{C4})$$

where statistical averaging is performed over a large ensemble of diffraction patterns.

Using the estimated value of  $\sigma_M^2$  [Eq. (C3)] we can determine a confidence interval for  $\langle A^2 \rangle_M$ , for a given  $\varepsilon$ . This can be done applying Tchebycheff's inequality<sup>41</sup>,

$$P \{ |\langle A^2 \rangle_M - \langle A^2 \rangle| < \varepsilon \} > 1 - \frac{\sigma_M^2}{\varepsilon^2}. \quad (\text{C5})$$

Eq. (C5) determines the probability  $P\{x\}$  that the average  $\langle A^2 \rangle_M$  lies in the interval  $\langle A^2 \rangle \pm \varepsilon$ .

As an example, we consider the system of  $N$  clusters with a uniform distribution of orientations  $\phi_k$  and determine the confidence interval for  $\langle A^2 \rangle_M$ , for a given accuracy  $\varepsilon$ . Taking into account that for such a system the probability density function  $p[A^2(m)]$  is known<sup>20</sup>

$$p[A^2(m)] = N e^{-N A^2(m)}, \quad 0 \leq A^2(m) \leq 1, \quad (\text{C6})$$

we obtain for the autocovariance (C4)

$$C(\Delta m) = \begin{cases} 1/N^2, & \Delta m = 0, \\ 0, & \Delta m \neq 0. \end{cases} \quad (\text{C7})$$

Substituting Eq. (C7) into Eq. (C3) we have for the variance

$$\sigma_M^2 = \frac{1}{M N^2}. \quad (\text{C8})$$

Using this value of  $\sigma_M^2$  we can determine the confidence interval for  $\langle A^2 \rangle_M$  according to Eq. (C5). For the 10% accuracy  $\varepsilon = 0.1\langle A^2 \rangle = 0.1/N$ , where we took into account that  $\langle A^2 \rangle = 1/N$  for a system considered here. With this we obtain from Tchebycheff inequality [Eq. (C5)]

$$P \{0.9\langle A^2 \rangle < \langle A^2 \rangle_M < 1.1\langle A^2 \rangle\} > 1 - \frac{10^2}{M}. \quad (\text{C9})$$

From this inequality the probability  $P$  that  $\langle A^2 \rangle_M$  lies in the interval from  $0.9\langle A^2 \rangle$  to  $1.1\langle A^2 \rangle$  is  $P > 0.9$  for  $M = 10^3$ , and  $P > 0.99$  for  $M = 10^4$  diffraction patterns.



TABLE I. Fourier components of the CCFs  $C_q^n$  ( $n = 10, 20$ ) for a single centered pentagonal cluster [ $C_q^n \text{ dilute} \equiv \mathcal{L}^n(q)$ ] and for a completely oriented dense system of clusters ( $C_q^n \text{ dense}$ ) described in Section IV (Fig. 3). The relative standard deviation is defined as  $\delta C_q^n = (C_q^n \text{ dense} - C_q^n \text{ dilute})/C_q^n \text{ dilute}$ .

Model	$C_q^{10}$					$C_q^{20}$
$q$ , [ $\text{nm}^{-1}$ ]	$q_1 = 0.023$	$q_2 = 0.029$	$q_3 = 0.036$	$q_4 = 0.043$	$q_5 = 0.059$	$q_5 = 0.059$
$C_q^n \text{ dilute} \equiv \mathcal{L}^n(q)$ , [Eq. (9)]	0.225	0.142	0.028	0.366	0.143	0.152
$C_q^n \text{ dense}$ , [Fig. 3(d)]	0.245	0.101	0.044	0.313	0.156	0.175
$\delta C_q^n$ , %	9	-29	57	-14	9	15

TABLE II. Fourier components of the CCFs  $C_q^n$  ( $n = 10, 20$ ) for a dilute ( $\langle C_q^n \text{ dilute} \rangle$ ) and dense ( $C_q^n \text{ dense}$ ) systems composed of 121 centered pentagonal clusters with the Gaussian distribution of orientations described in Section IV (Fig. 4). The parameter  $\delta C_q^n$  is defined as in Table I.

Model	$C_q^{10}$					$C_q^{20}$
$q$ , [ $\text{nm}^{-1}$ ]	$q_1 = 0.023$	$q_2 = 0.029$	$q_3 = 0.036$	$q_4 = 0.043$	$q_5 = 0.059$	$q_5 = 0.059$
$\langle C_q^n \text{ dilute} \rangle$ , [Eq. (12)]	0.152	0.096	0.019	0.247	0.096	0.031
$C_q^n \text{ dense}$ , [Fig. 4(d)]	0.263	0.078	0.016	0.233	0.071	0.024
$\delta C_q^n$ , %	73	-18	-15	-6	-26	-23

TABLE III. Fourier components of the averaged CCFs  $\langle C_q^n \rangle_M$  ( $n = 10, 20$ ) for a dilute ( $\langle C_q^{n \text{ dilute}} \rangle_M$ ,  $M \rightarrow \infty$ ) and dense ( $\langle C_q^{n \text{ dense}} \rangle_M$ ) systems composed of  $N = 11$  centered pentagonal clusters with the uniform distribution of orientations. The parameter  $\delta C_q^n$  is defined as in Table I.

Model	$\langle C_q^{10} \rangle_M$					$\langle C_q^{20} \rangle_M$
$q$ , [nm <sup>-1</sup> ]	$q_1 = 0.023$	$q_2 = 0.029$	$q_3 = 0.036$	$q_4 = 0.043$	$q_5 = 0.059$	$q_5 = 0.059$
$\langle C_q^{n \text{ dilute}} \rangle_M$ , [Eqs. (22)]	2.05E-2	1.29E-2	0.25E-2	3.33E-2	1.30E-2	1.38E-2
$\langle C_q^{n \text{ dense}} \rangle_M$ , [Fig. 7(a)]	2.56E-2	1.67E-2	0.56E-2	3.56E-2	1.37E-2	1.49E-2
$\delta C_q^n$ , %	25	29	120	7	15	13

TABLE IV. Fourier components of the CCFs  $C_q^n$  ( $n = 10, 20$ ) for the case of a fully coherent scattering [ $C_q^{n \text{ coh}}$ ] and a partially coherent scattering ( $C_q^{n \text{ pcoh}}$ ,  $l_{\text{coh}} = 600$  nm) from a system of 11 clusters with uniform distribution of orientations. The relative standard deviation is defined as  $\delta C_q^n = [C_q^{n \text{ pcoh}} - C_q^{n \text{ coh}}]/C_q^{n \text{ coh}}$ .

Model	$C_q^{10}$					$C_q^{20}$
$q$ , [nm <sup>-1</sup> ]	$q_1 = 0.023$	$q_2 = 0.029$	$q_3 = 0.036$	$q_4 = 0.043$	$q_5 = 0.059$	$q_5 = 0.059$
$C_q^{n \text{ coh}}$ , [Fig. 6(c)]	0.077	0.050	0.024	0.107	0.045	0.022
$C_q^{n \text{ pcoh}}$ , [Fig. 8(b)]	0.050	0.040	0.008	0.063	0.031	0.019
$\delta C_q^n$ , %	-35	-20	-68	-42	-31	-12

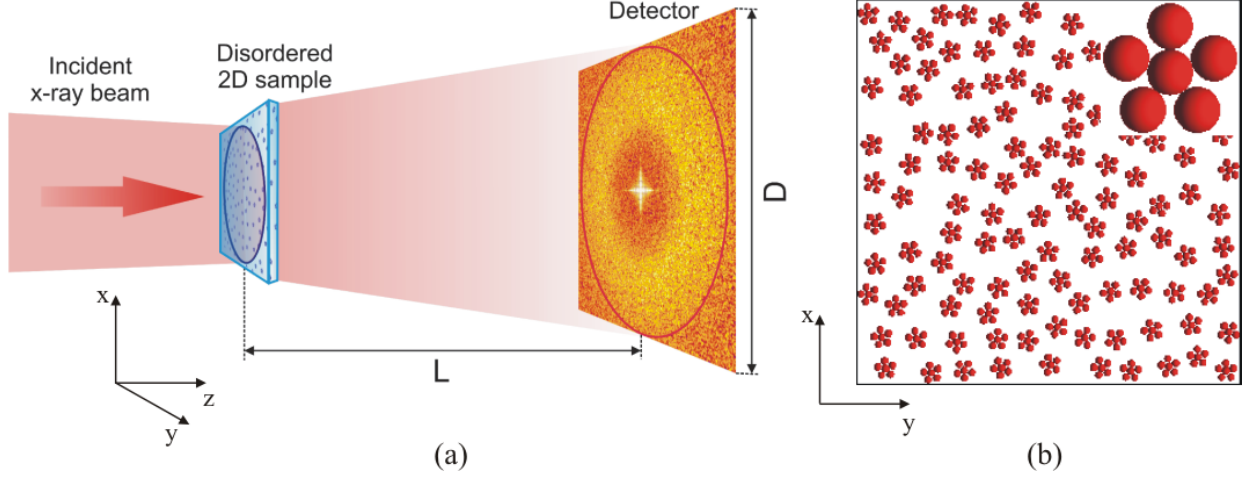


FIG. 1. (Color online) (a) Geometry of the diffraction experiment. A coherent x-ray beam illuminates a 2D disordered sample and produces a diffraction pattern on a detector. The direction of the incident beam is defined along the  $z$  axis of the coordinate system. (b) A disordered 2D sample composed of centered pentagonal clusters (an enlarged view of the cluster is shown in the inset). The case when all clusters have random position and orientation in the 2D plane is shown.

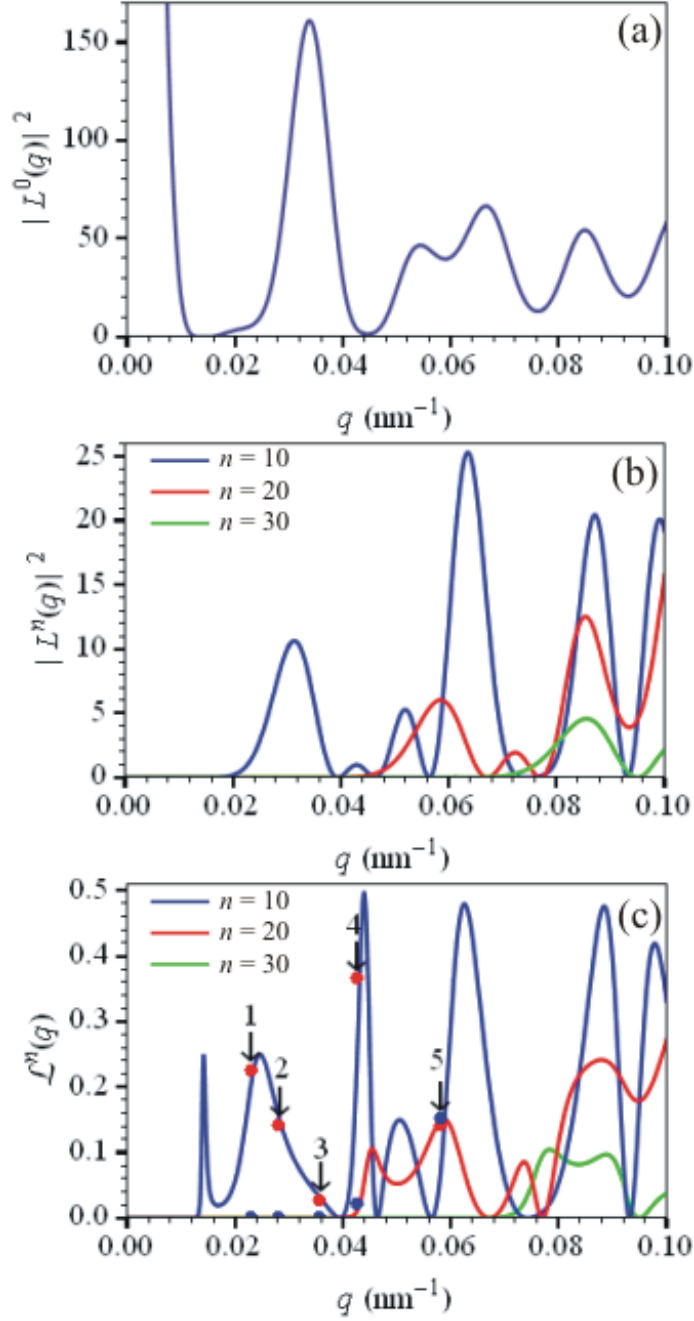


FIG. 2. (Color online) The structural terms  $|L^n(q)|^2$  [see Eqs. (7) and (A3)] calculated for a single centered pentagonal cluster as a function of  $q$  for (a)  $n = 0$  and (b)  $n = 10, 20$  and  $30$ . These terms are normalized by the value of the form-factor  $|f(q)|^2$  of a sphere. (c) The normalized values  $\mathcal{L}^n(q) = |L^n(q)/L^0(q)|^2$  of the same Fourier components. Arrows indicate the  $q$  values at which the CCFs presented in this paper were calculated (circles mark the values of the specific Fourier components at these positions).

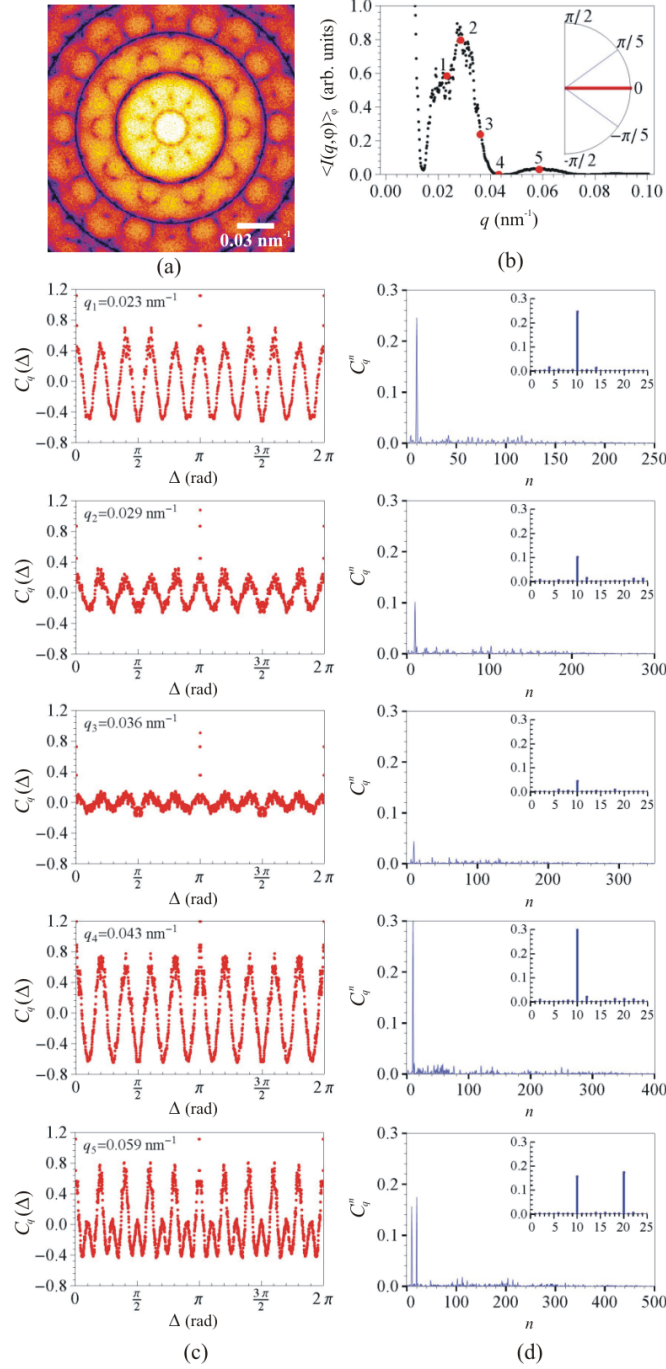


FIG. 3. (Color online) Cross-correlation analysis of the diffraction data for a completely oriented dense system ( $\langle R \rangle/d \approx 1.5$ ) consisting of 121 pentagonal clusters. (a) Diffraction pattern; (b) Angular averaged intensity  $\langle I(q, \varphi) \rangle_\varphi$ . (inset) Angular distribution of clusters in the sample. In this case all clusters have the same orientation; (c) CCFs  $C_q(\Delta)$  calculated at the selected  $q$  values, indicated in (b); (d) Fourier spectra  $C_q^n$  of the CCFs for each selected  $q$  value. (insets) Enlarged view of the first 25 Fourier components.

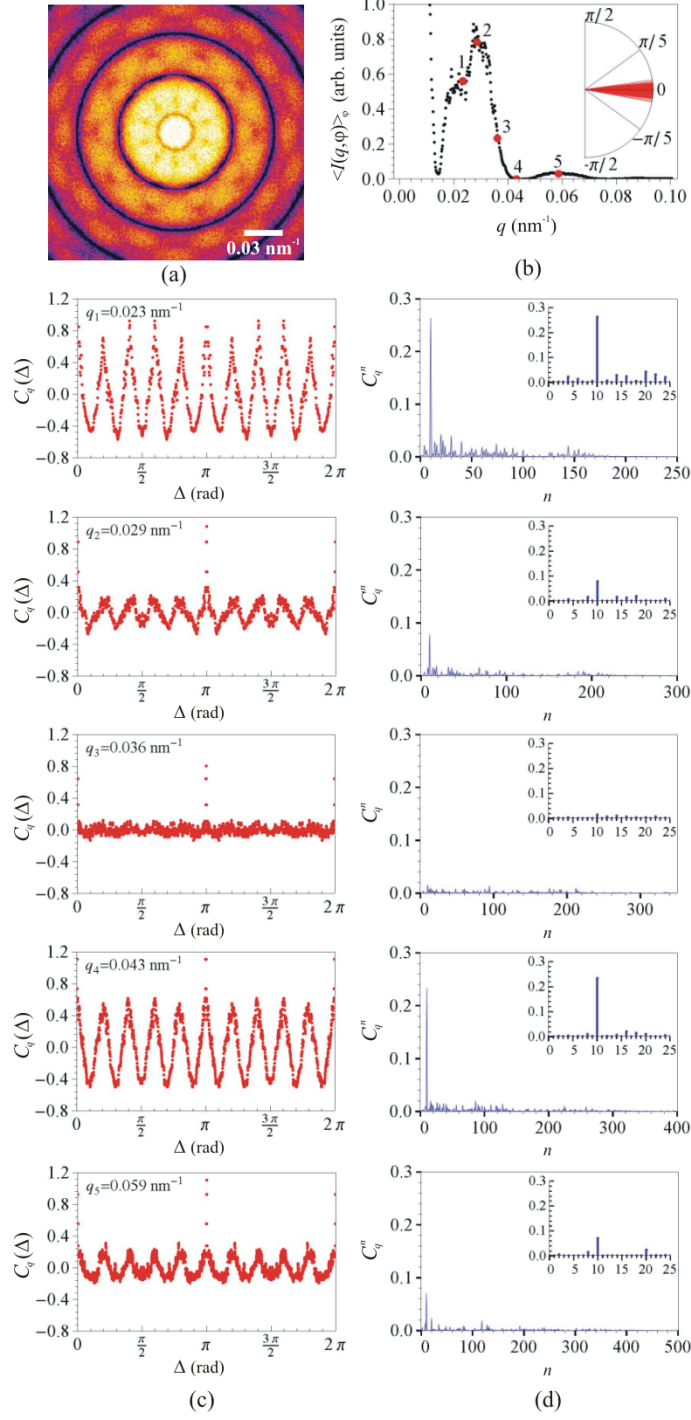


FIG. 4. (Color online) Cross-correlation analysis of the diffraction data (similar to Fig. 3) for a system consisting of 121 pentagonal clusters described by the Gaussian distribution of cluster orientations with the standard deviation  $\sigma_\phi = 0.05 \cdot 360^\circ / 5 = 3.6^\circ$  and zero mean [see inset in (b)].

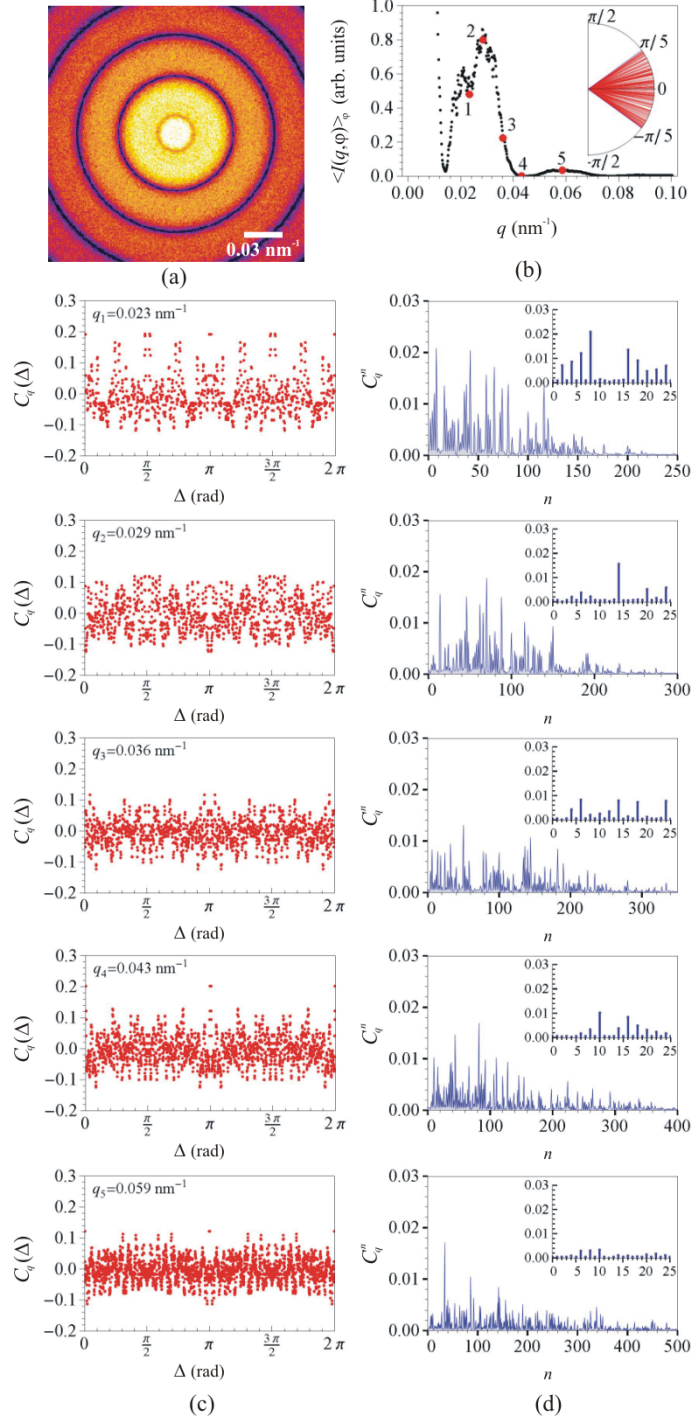


FIG. 5. (Color online) Cross-correlation analysis of the diffraction data (similar to Figs. 3 and 4) for a system consisting of 121 pentagonal clusters described by the uniform distribution of cluster orientations in the angular range  $(-\pi/5, \pi/5)$  [see inset in (b)].

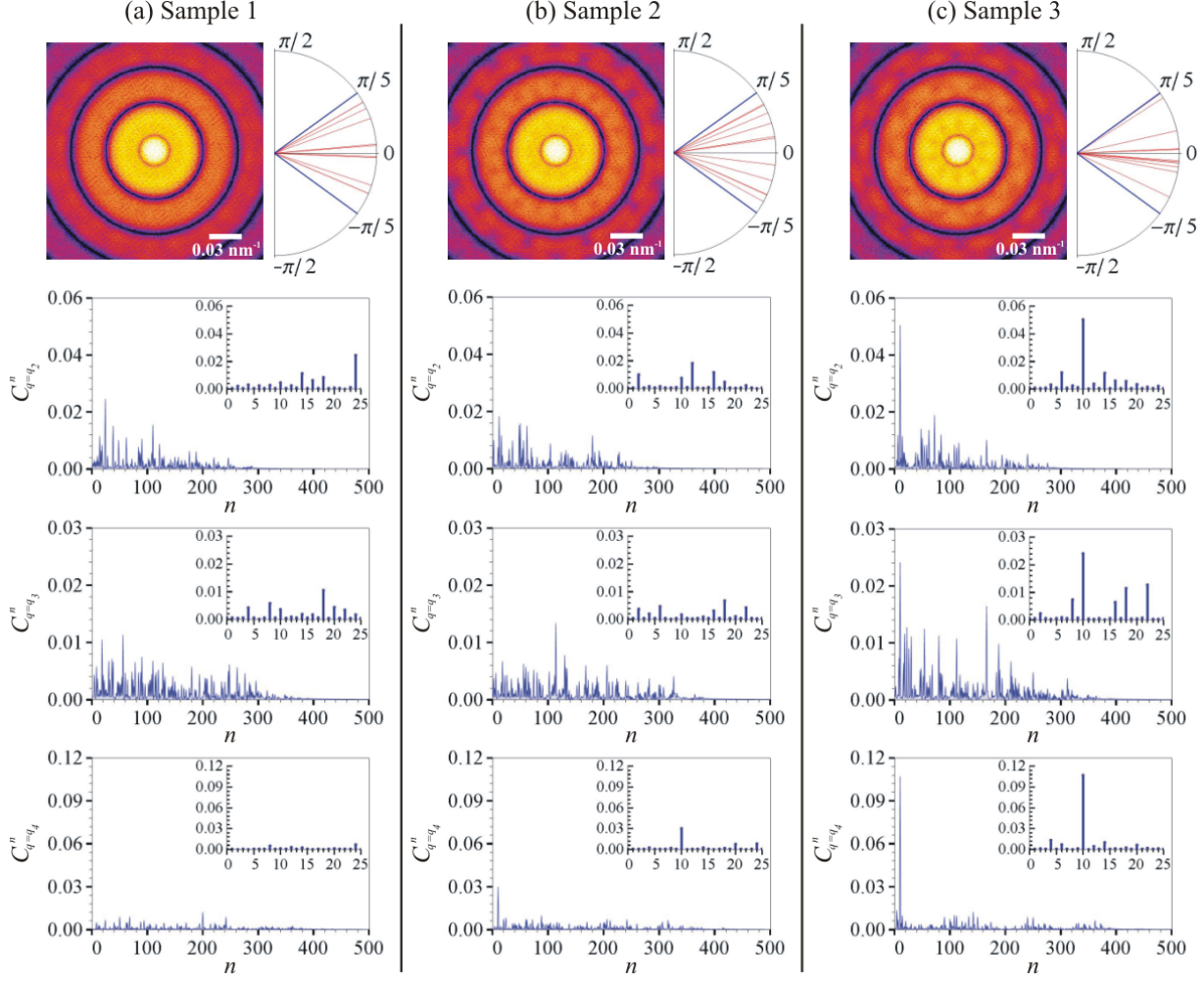


FIG. 6. (Color online) Diffraction patterns and Fourier components of the CCFs calculated for three disordered systems, containing 11 clusters in different, uniformly distributed orientations. Angular distribution of clusters in each sample is given in the angular diagrams. Fourier spectra were calculated for each system at three  $q$  values,  $q_2 = 0.029 \text{ nm}^{-1}$ ,  $q_3 = 0.036 \text{ nm}^{-1}$ , and  $q_4 = 0.043 \text{ nm}^{-1}$ .



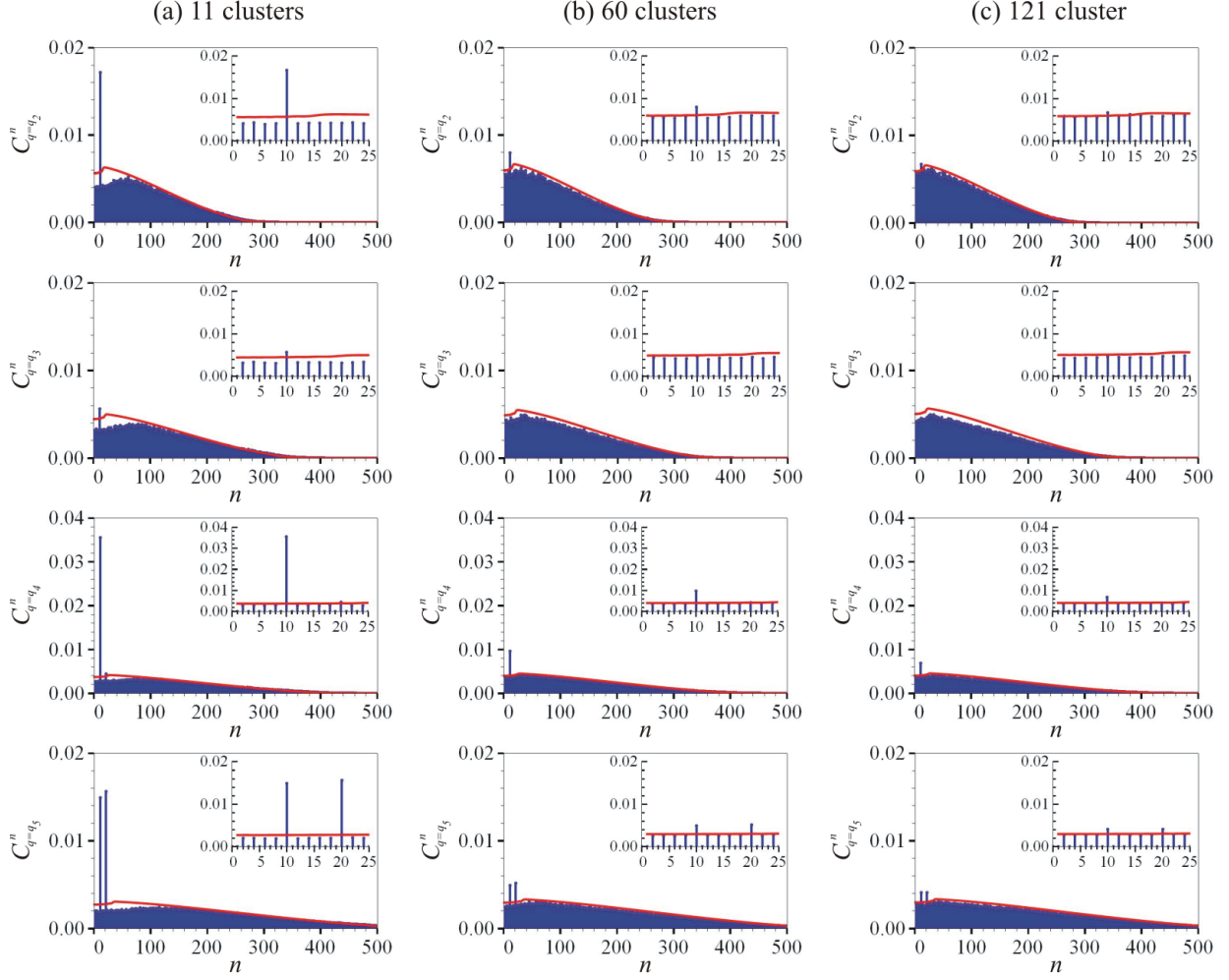


FIG. 7. (Color online) Fourier spectra of the CCFs  $\langle C_q(\Delta) \rangle_M$ , averaged over  $M = 1000$  diffraction patterns. Three disordered systems composed of (a) 11, (b) 60 and (c) 121 clusters with the uniform distribution of angular orientations are considered. Fourier spectra were calculated for each system at four  $q$  values,  $q_2 = 0.029 \text{ nm}^{-1}$ ,  $q_3 = 0.036 \text{ nm}^{-1}$ ,  $q_4 = 0.043 \text{ nm}^{-1}$  and  $q_5 = 0.059 \text{ nm}^{-1}$ . Red solid line in each Fourier spectrum defines the magnitude of the inter-cluster contribution estimated using Eq. (B10).

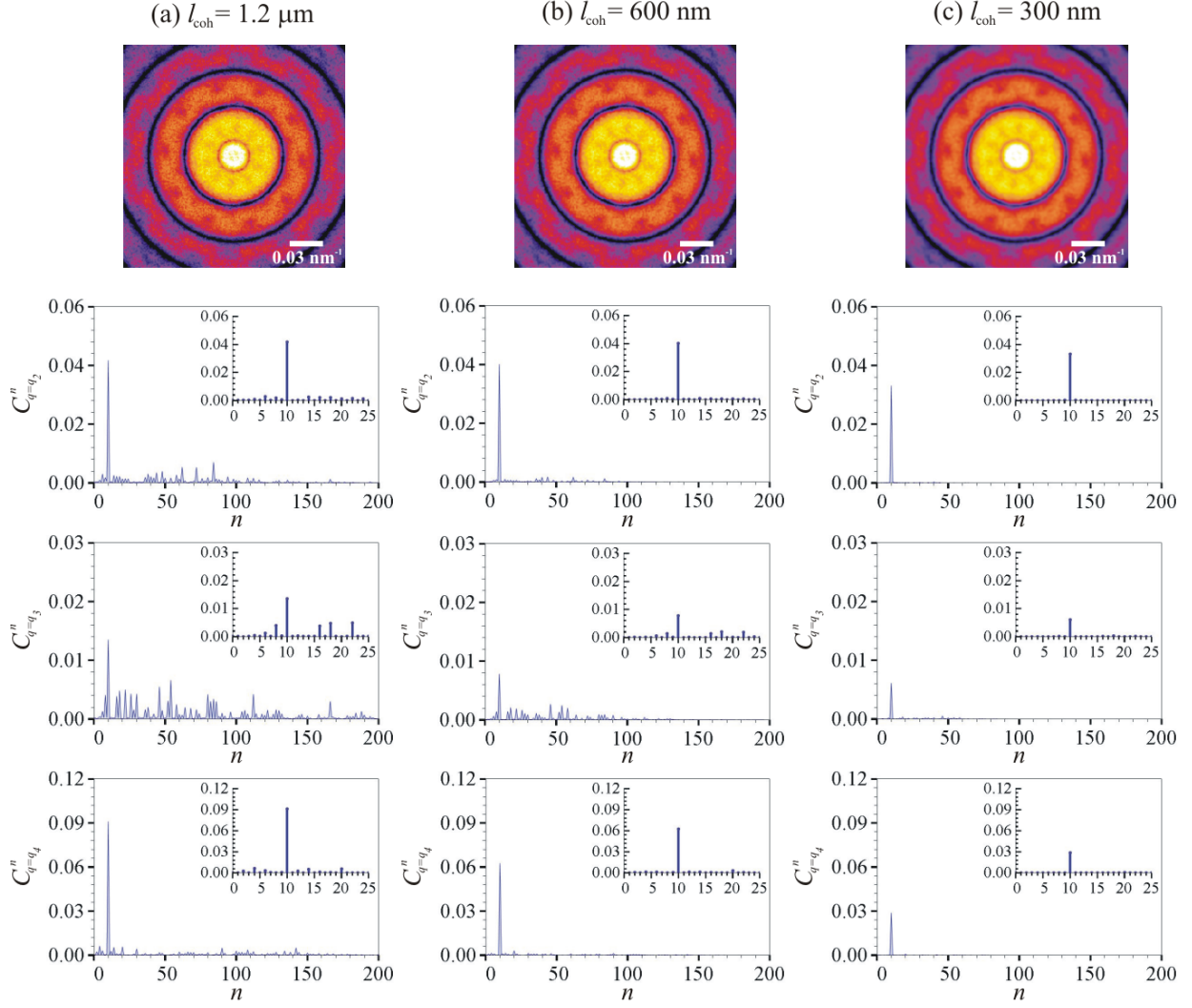


FIG. 8. (Color online) Diffraction patterns and Fourier components of the CCFs calculated for a disordered system consisting of 11 clusters in the same positions and orientations as in the sample 3 in Fig. 6. Different values of the transverse coherence length  $l_{\text{coh}}$  of the incoming x-ray beam: (a)  $l_{\text{coh}} = 1.2 \mu\text{m}$ , (b)  $l_{\text{coh}} = 600 \text{ nm}$ , and (c)  $l_{\text{coh}} = 300 \text{ nm}$  were considered. Fourier spectra were calculated for each case at three  $q$  values,  $q_2 = 0.029 \text{ nm}^{-1}$ ,  $q_3 = 0.036 \text{ nm}^{-1}$ , and  $q_4 = 0.043 \text{ nm}^{-1}$ .

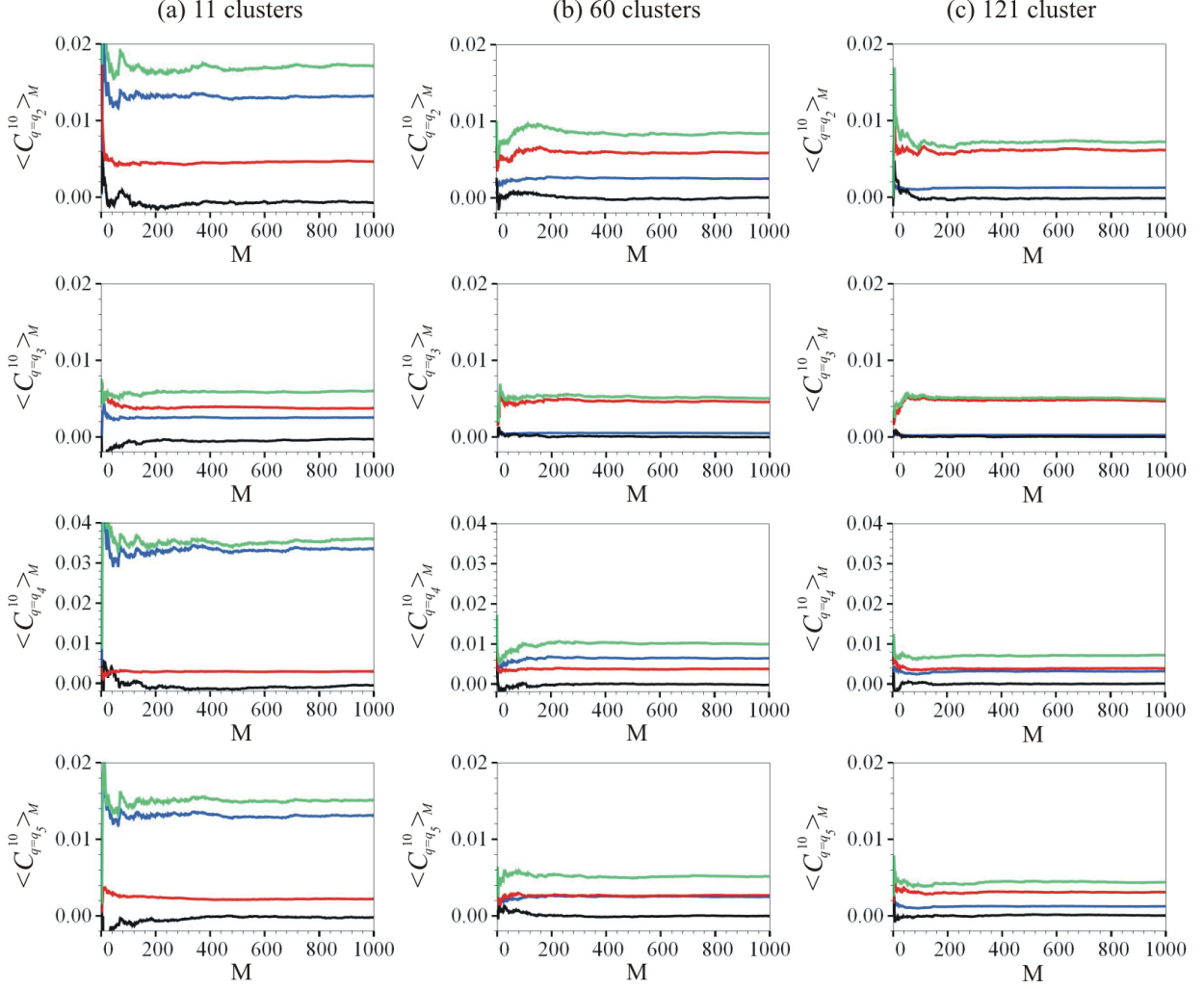


FIG. 9. (Color online) (Color online) The evolution of different terms in the expansion of the Fourier component  $\langle C_q^n \rangle_M$  ( $n = 10$ ) [Eq. (19)] as a function of the number  $M$  of diffraction patterns used in the averaging. The results are presented for the same  $q$  values as in Fig. 7. The following contributions are shown:  $\langle S_1^n \rangle_M / \langle |I^0(q)|^2 \rangle_M$  (blue curve),  $\langle S_2^n + S_3^n \rangle_M / \langle |I^0(q)|^2 \rangle_M$  (black curve),  $\langle S_4^n \rangle_M / \langle |I^0(q)|^2 \rangle_M$  (red curve), and the sum of all terms  $\langle C_q^n \rangle_M$  (green curve). Three disordered systems composed of (a) 11, (b) 60 and (c) 121 clusters with the uniform distribution of angular orientations are considered.

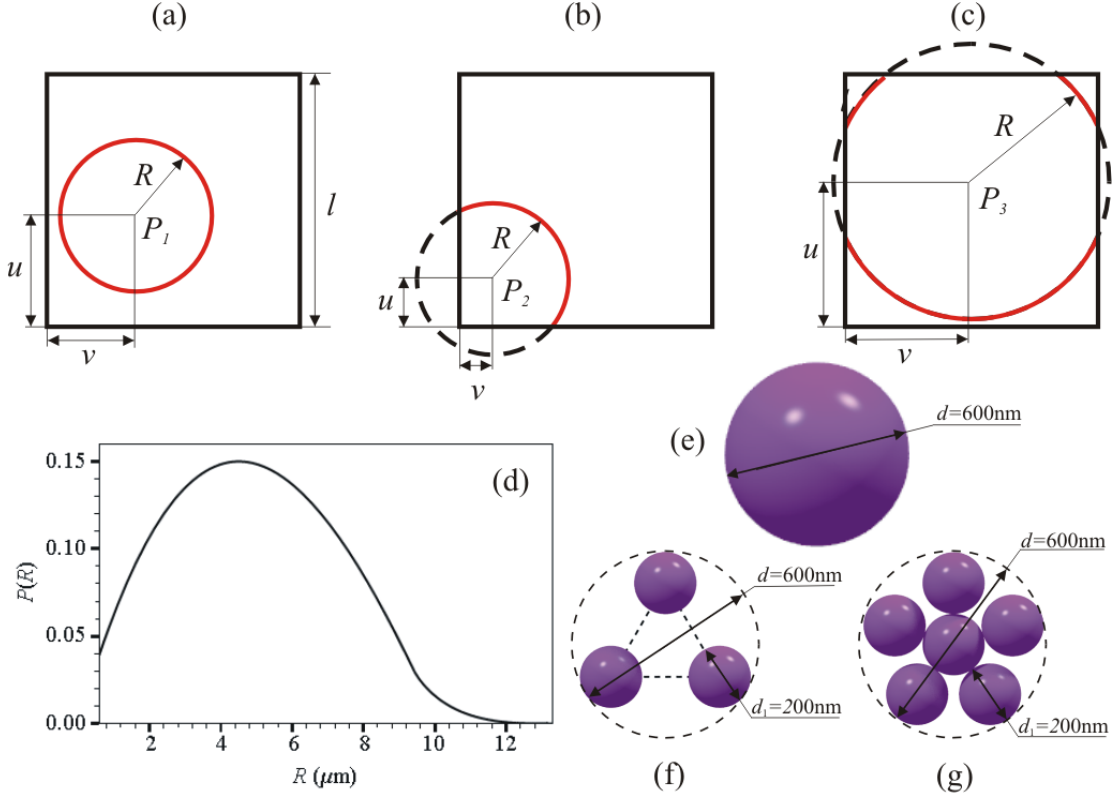


FIG. 10. (Color online) Derivation of the RDF for a square sample. (a)-(c) Three cases when the reference particle is placed at the position  $P_i$  ( $i = 1, 2, 3$ ), which is located from the sides of the square at distances  $u$  and  $v$ , are shown.  $P_{\text{part}}(R, u, v)$  is proportional to (a) the length of the circumference of the radius  $R$ , (b) the length of the single arc bounded by the square, and (c) the total length of three arcs bounded by the square (in all cases solid red parts of the circumference); (d) RDF for spherical particles of a diameter  $d = 600 \text{ nm}$  distributed within the square 2D sample with the length of a side of the square  $l = 10 \mu\text{m}$ ; Three model systems used in calculations of inter-particle contribution consist of (e) spherical particles (f) triangular clusters and (g) centered pentagonal clusters.

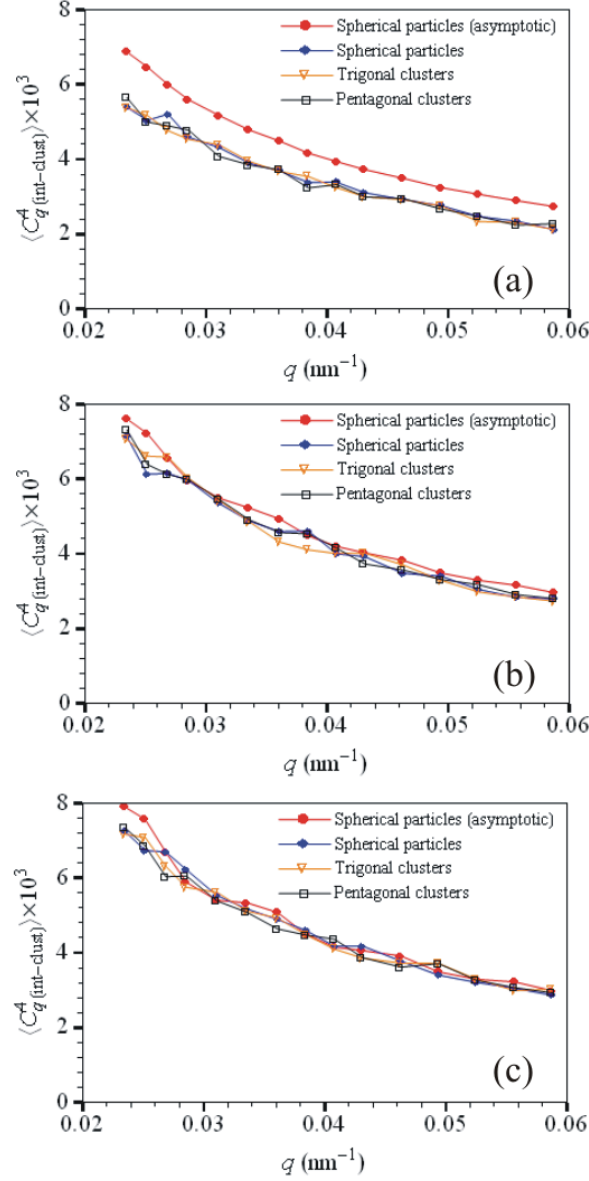


FIG. 11. (Color online) Fourier components of the average CCFs corresponding to the inter-particle correlations  $\langle C_q^n(\text{int-clust}) \rangle$  (for  $n = 4$ ). Calculations were performed using asymptotic expressions [Eqs. (B10), (B16a) and (B16b)] for a system composed of spherical particles (red circles). For comparison, the results of direct averaging of Eq. (B1) over the set of  $M = 1000$  diffraction patterns for a system composed of spherical particles (blue diamonds), trigonal clusters (yellow triangles) and pentagonal clusters (black squares) are shown. The number of particles in the system (a)  $N = 11$ , (b)  $N = 60$ , and (c)  $N = 121$ .